Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON,
University of California, Los Angeles
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY,
University of California, Los Angeles
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

ARTHUR C. HOFFMAN, Production Manager
HELEN ORR, Promotion Manager

Volume 62, 1961

BUREAU OF EDNL. & PSYCL. RESEARCH
LIBRARY

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

Second-class postage paid at Lancaster, Pa.

CONTENTS OF VOLUME 62

Archer, E. J. **Supplementary Report: Familiarity or Low Validity?** 207

Archer, E. J. See Erlebacher, A.

Arnoult, M. D., and Price, C. W. **Pattern Matching in the Presence of Visual Noise** 72

Atkinson, R. C. **The Observing Response in Discrimination Learning** 253

Barnes, G. W., and Kish, G. B. **Reinforcing Properties of the Onset of Auditory Stimulation** 164

Bastian, J. **Associative Factors in Verbal Transfer** 70

Berlyne, D. E. **Conflict and the Orientation Reaction** 476

Bialer, I. **Primary and Secondary Stimulus Generalization as Related to Intelligence Level** 395

Birch, D. See Ison, J. R.

Birch, D., Clifford, L. T., and Butterfield, J. **Response Latency as a Function of Size of Gap in the Elevated Runway** 179

Bower, G. H. **A Contrast Effect in Differential Conditioning** 196

Brackett, H. R. See Reid, L. S.

Braden, I. See Jones, A.

Brown, L. T. See Gagné, R. M.

Brownstein, A. J. See Marx, M. H.

Butterfield, J. See Birch, D.

Campbell, D. T. See Krantz, D. L.

Carterette, T. S. **An Application of Stimulus Sampling Theory to Summated Generalization** 448

Champion, R. A. **Supplementary Report: Interpolated UCS Trials in GSR Conditioning** 206

Champion, R. A., and Jones, J. E. **Forward, Backward, and Pseudoconditioning of the GSR** 58

Chapman, J. P. **The Spacing of Sequentially Dependent Trials in Probability Learning** 545

Cicala, G. A. **Running Speed in Rats as a Function of Drive Level and Presence or Absence of Competing Response Trials** 329

Clifford, L. T. See Birch, D.

Coke, E. U. See Rothkopf, E. Z.

Collier, G., Knarr, F. A., and Marx, M. H. **Some Relations between the Intensive Properties of the Consummatory Response and Reinforcement** 484

Collier, G., and Willis, F. N. **Deprivation and Reinforcement** 377

Di Vesta, F. J. **Contrast Effects in the Verbal Conditioning of Meaning** 535

Drazin, D. H. **Effects of Foreperiod, Foreperiod Variability, and Probability of Stimulus Occurrence on Simple Reaction Time** 43

Duncan, C. P. See Lewis, D. J.

Edelberg, R. **The Relationship between the Galvanic Skin Response, Vasoconstriction, and Tactile Sensitivity** 187

Edwards, W. **Probability Learning in 1000 Trials** 385

Edwards, W. See Lindman, H.

Erlebacher, A., and Archer, E. J. **Perseveration as a Function of Degree of Learning and Percentage of Reinforcement in Card Sorting** 510

Erlick, D. E. **Judgments of the Relative Frequency of a Sequential Series of Two Events** 105

Eijkman, E., and Vendrik, A. J. H. **Dynamic Behavior of the Warmth Sense Organ** 403

Feldman, S. M. **Differential Effects of Shock in Human Maze Learning** 171

Fivel, M. W. See Osler, S. F.

Forsythe, J. B. See Gardner, R. A.

Gagné, R. M., and Brown, L. T. Some Factors in the Programing of Conceptual Learning ... 313
Gannon, D. R., and Noble, C. E. Familiarization (*n*) as a Stimulus Factor in Paired-Associate Verbal Learning ... 14
Gardner, R. A., and Forsythe, J. B. Supplementary Report: Two-Choice Decision Behavior with Many Alternative Events ... 631
Goldstein, A. G. Familiarity and Apparent Complexity of Random Shapes ... 594
Goldstein, H. See Spence, K. W.
Gorman, J. See King, M. S.
Gormezano, I. See Moore, J. W.
Goss, A. E. See Levitt, H.
Goss, A. E., and Sugerman, M. E. Paired-Associates Learning with Varying Relative Percentages of Occurrence of Alternative Response Members ... 24
Green, B. F., Jr. Figure Coherence in the Kinetic Depth Effect ... 272
Gregg, L. W. See Karn, H. W.
Gutman, N. J. See Kamen, J. M.
Hanawalt, N. G., and Tarr, A. G. The Effect of Recall upon Recognition ... 361
Hawkins, W. F. See Reid, L. S.
Hickson, R. H. Response Probability in a Two-Choice Learning Situation with Varying Probability of Reinforcement ... 138
Hilgard, E. R. See Umemoto, T.
Hirsh, I. J., and Sherrick, C. E., Jr. Perceived Order in Different Sense Modalities 423
Hogan, J. A. Copying Redundant Messages ... 153
Horowitz, L. M. Free Recall and Ordering of Trigrams ... 51
Hunt, E. B. Memory Effects in Concept Learning ... 598
Ison, J. R., and Birch, D. T-Maze Reversal Following Differential Endbox Placement ... 200
Jahnke, J. C. Postrest Motor Learning Performance as a Function of Degree of Learning ... 605
Jakobovits, L. A., and Lambert, W. E. Semantic Satiation among Bilinguals ... 576
Jensen, G. D. Partial Reinforcement Effects (PREs) and Inverse PREs Determined by Position of a Nonrewarded Block of Responses ... 461
Jones, A. The Relative Effectiveness of Positive and Negative Verbal Reinforcers 368
Jones, A. Supplementary Report: Information Deprivation and Irrelevant Drive as Determiners of an Instrumental Response ... 310
Jones, A., Wilkinson, H. J., and Braden, I. Information Deprivation as a Motivational Variable ... 126
Jones, J. E. The CS–UCS Interval in Conditioning Short- and Long-Latency Responses ... 612
Jones, J. E. See Champion, R. A.
Kamen, J. M., Pilgrim, F. J., Gutman, N. J., and Kroll, B. J. Interactions of Suprathreshold Taste Stimuli ... 348
Kanareff, V. T. See Lanzetta, J. T.
Kanfer, F. H. See Smock, C. D.
Karn, H. W., and Gregg, L. W. Acquisition of Perceptual Responses as a Function of Loading, Location, and Repetition ... 62
Karn, H. W., and Gregg, L. W. Supplementary Report: Effects of Instructions on the Perception of Multiple Targets ... 533
Kimble, G. A. See King, M. S.
King, M. S., Kimble, G. A., Gorman, J., and King, R. A. Replication Report: Two Failures to Reproduce Effects of Anxiety on Eyelid Conditioning ... 532
King, R. A. See King, M. S.
Kish, G. B. See Barnes, G. W.
Koehler, J., Jr. Role of Instructions in Two-Choice Verbal Conditioning with Contingent Partial Reinforcement ... 122
Knarr, F. A. See Collier, G.
Krantz, D. L., and Campbell, D. T. Separating Perceptual and Linguistic Effects of Context Shifts upon Absolute Judgments ... 35
Kroll, B. J. See Kamen, J. M.

LABERGE, D. Generalization Gradients in a Discrimination Situation.............. 88
LACEY, H. M. Mediating Verbal Responses and Stimulus Similarity as Factors in Conceptual Naming by School Age Children.............................. 113
LACHMAN, R. The Influence of Thirst and Schedules of Reinforcement-Nonreinforcement Ratios upon Brightness Discrimination............................ 80
LAMBERT, W. E. See JAKOBOVITS, L. A.
LANZETTA, J. T., AND KANAREFF, V. T. The Effects of Congruent and Conflicting Social and Task Feedback on the Acquisition of an Imitative Response........ 322
LEVITT, H., AND GOSS, A. E. Stimulus Attributes and Drive in Paired-Associate Learning.. 243
LEWIS, D. J., AND DUNCAN, C. P. Effects of Variable Magnitude of Reward on a Lever-Pulling Response.. 203
LINDMAN, H., AND EDWARDS, W. Supplementary Report: Unlearning the Gambler's Fallacy.. 630
LIPMAN, R. S., AND SPITZ, H. H. The Relationship between Kinesthetic Satiation and Inhibition in Rotary Pursuit Performance................................ 468
LLOYD, K. E. Supplementary Report: Short-Term Retention as a Function of Average Storage Load... 632
LLOYD, K. E. See REID, L. S.
LORDAHL, D. S. Concept Identification Using Simultaneous Auditory and Visual Signals.. 283
MCGEHEE, N. E., AND SCHULZ, R. W. Mediation in Paired-Associate Learning..... 565
MCGUIRE, W. J. A Multiprocess Model for Paired-Associate Learning............ 335
MARTIN, E. J. See THOMPSON, M. E.
MARX, M. H. See COLLIER, G.
MARX, M. H., MURPHY, W. W., AND BROWNSTEIN, A. J. Recognition of Complex Visual Stimuli as a Function of Training with Abstracted Patterns........... 456
MOORE, J. W., AND GORMEZANO, I. Yoked Comparisons of Instrumental and Classical Eyelid Conditioning.. 552
MURDOCK, B. B., JR. The Retention of Individual Items......................... 618
MURPHY, W. W. See MARX, M. H.
MYERS, J. L., REILLY, R. E., AND TAUB, H. A. Differential Cost, Gain, and Relative Frequency of Reward in a Sequential Choice Situation....................... 357
NOBLE, C. E. See GANNON, D. R.
OSLER, S. F., AND FIVEL, M. W. Concept Attainment: I. The Role of Age and Intelligence in Concept Attainment by Induction.............................. 1
OSLER, S. F., AND TRAUTMAN, G. E. Concept Attainment: II. Effect of Stimulus Complexity upon Concept Attainment at Two Levels of Intelligence.......... 9
PILGRIM, F. J. See KAMEN, J. M.
PINNEO, L. R. The Effects of Induced Muscle Tension during Tracking on Level of Activation and on Performance.. 523
PRICE, C. W. See ARNOULT, M. D.
PROKASY, W. F., AND WHALEY, F. L. The Intertrial Interval in Classical Conditioning... 560
REID, L. S., LLOYD, K. E., BRACKETT, H. R., AND HAWKINS, W. F. Short-Term Retention as a Function of Average Storage Load and Average Load Reduction.. 518
REILLY, R. E. See MYERS, J. L.
RETHLINGSHAFER, D., AND SHERRER, T. I. Supplementary Report: Effect of Practice on an Illusion.. 95
RIACH, W. J. See SALTZ, E.
ROTHKOPF, E. Z., AND COKE, E. U. The Prediction of Free Recall from Word Association Measures... 433
RUBINSTEIN, I. Supplementary Report: The Influence of One Stimulus on the Prediction of the Alternative Stimulus in Two-Choice Problems.................. 311
SALTZ, E. Response Pretraining: Differentiation or Availability?................ 583
SALTZ, E., AND RIACH, W. J. The Effect of Stress on Stimulus Differentiation..... 588
SCHULZ, R. W. See MCGEHEE, N. E.
SCHULZ, R. W. See UNDERWOOD, B. J.
SHELLY, M. W. Learning with Reduced Feedback Information...................... 209

Shepard, R. N., and Teghtsoonian, M. Retention of Information under Conditions Approaching a Steady State........ 302
Sherrer, T. L. See Rethlingshafer, D.
Sherrick, C. E., Jr. See Hirsh, I. J.
Shinkman, P. G. Supplementary Report: Perception of Stimuli of Varying Dimensionality........ 626
Simpson, W., and Voss, J. F. Psychophysical Judgments of Probabilistic Stimulus Sequences........ 416
Slamecka, N. J. Proactive Inhibition of Connected Discourse........ 295
Smith, O. W. See Smith, P. C.
Smith, P. C., and Smith, O. W. Ball Throwing Responses to Photographically Portrayed Targets........ 223
Smith, P. C., and Smith, O. W. Veridical Perceptions of Cylindricality: A Problem of Depth Discrimination and Object Identification........ 145
Smock, C. D., and Kanfer, F. H. Response Bias and Perception........ 158
Spence, K. W., and Goldstein, H. Eyelid Conditioning Performance as a Function of Emotion-Producing Instructions........ 291
Spitz, H. H. See Lipman, R. S.
Sugerman, M. E. See Goss, A. E.
Tarr, A. G. See Hanawalt, N. G.
Taub, H. A. See Myers, J. L.
Teghtsoonian, M. See Shepard, R. N.
Thompson, M. E., and Martin, E. J. Supplementary Report: Alternation in a T Maze with Habit Held Constant........ 629
Trautman, G. E. See Osler, S. F.
Umemoto, T., and Hilgard, E. R. Paired-Associate Learning as a Function of Similarity: Common Stimulus and Response Items within the List........ 97
Underwood, B. J., and Schulz, R. W. Studies of Distributed Practice: XXI. Effect of Interference from Language Habits........ 571
Vendrik, A. J. H. See Eijkman, E.
Voss, J. F. See Simpson, W.
Wagner, A. R. Effects of Amount and Percentage of Reinforcement and Number of Acquisition Trials on Conditioning and Extinction........ 234
Waugh, N. C. Free Versus Serial Recall........ 496
Whaley, F. L. See Prokasy, W. F.
Wilkinson, H. J. See Jones, A.
Wilkinson, R. T. Interaction of Lack of Sleep with Knowledge of Results, Repeated Testing, and Individual Differences........ 263
Williams, J. P. Supplementary Report: A Selection Artifact in Rock's Study of the Role of Repetition........ 627
Willis, F. N. See Collier, G.
Witte, R. S. Conditional Response Probability in a T Maze........ 439
Wunderlich, R. A. Strength of a Generalized Conditioned Reinforcer as a Function of Variability of Reward........ 409
Zipf, S. G. The Effects of Amount of Reward, Requirement, and Several Related Probabilities on Human Performance........ 503

Journal of
Experimental Psychology

VOL. 62, NO. 1 JULY 1961

CONCEPT ATTAINMENT: I. THE ROLE OF AGE AND INTELLIGENCE IN CONCEPT ATTAINMENT BY INDUCTION[1]

SONIA F. OSLER AND MYRNA WEISS FIVEL[2]

Goucher College

Department of Pediatrics, Johns Hopkins University

Most concept studies with children have utilized either the interview-questionnaire method or the performance method (Vinacke, 1952, p. 108). Each of these procedures presents certain obstacles to the discovery of *S*'s knowledge of a specific concept or his ability to utilize it. Implicit in the interview-questionnaire method, illustrated in the early works of Piaget (1929, 1930) and others (Klingberg, 1957; Russel, 1940, 1942), is the assumption that knowledge of a concept is necessarily accompanied by the child's ability to translate it into words. That this assumption is not justified has been demonstrated in a variety of ways with adults and children (Hazlitt, 1930; Heidbreder, 1927; Hull, 1920). It is primarily for this reason that Piaget modified his early procedures involving only interviews to the combined interview performance technique.

With few exceptions the performance procedures utilized in concept studies boil down essentially to a type of sorting technique. The tasks may vary from matching to stimulus (Brian & Goodenough, 1929; Colby & Robertson, 1942) to sorting into categories that "go together" (Thompson, 1941; Welch, 1940). The difficulty associated with this technique is that performance may be a function

---

[1] This investigation was supported by a grant from the National Institute of Mental Health, Psychopharmacology Service Center, made to Leon Eisenberg of Johns Hopkins University, and the senior author (Grant MY-2583).

The writers wish to express their gratitude to Mary A. Adams, Assistant Superintendent, Elementary Education, Lorne H. Woollatt, Former Director of Bureau of Research, Vernon S. Vavrina, Director, Junior and Senior High Schools, and to Arthur Lichtenstein, Director of Special Services for Pupils, Baltimore City Department of Education; and to B. Melvin Cole, Director, Elementary Education, Norris A. King, Director, Secondary Education, and Charles J. Leiman, Supervisor, Psychological Services, Baltimore County Department of Education; and to the the principals and their staffs who cooperated in the conduct of the study.

They also wish to thank Grace E. Trautman, Nancy A. Freydberg, and Sandra J. Weiss for testing the subjects and for many suggestions concerning the conduct of the study; and to Natalia Chapanis for assistance with the statistical analysis of the data. To Jerome Cornfield and Margaret P. Martin go our special thanks for their invaluable assistance with the statistical problems encountered in this investigation.

[2] Now at the University of Pennsylvania.

of factors other than knowledge of the specific concept under investigation. For example, Heidbreder (1947) has demonstrated with adult *S*s a dominance hierarchy in concept learning and sorting, in which object, form, and number concepts occupy in this order decreasing positions in the hierarchy; and Brian & Goodenough (1929) have similarly demonstrated with children the relative dominance of color and form. Others have shown that specific stimulus attributes may affect their relative dominance (Dattman & Israel, 1951; Heidbreder, 1952; Stevenson & McBee, 1958). Furthermore, it is reasonable to assume that where alternative possibilities for categorizing exist the preference, interest, or experience of *S* will affect his performance.

It would, therefore, seem desirable to use a technique for concept studies which would avoid the difficulties associated with the interview-questionnaire and sorting techniques and yield a sensitive measure of *S*'s capacity to deal with the task. The inductive method seems to fulfill these requirements. In this method *S* is required to give the same response to differing stimuli which belong to a common category. The *S* demonstrates his attainment of the concept when he can respond appropriately to new stimuli at a level of accuracy specified by *E*. The stimuli are so designed that in order to achieve the required level of success *S* must categorize according to the concept expected by *E*. In this manner competition from other concepts is diminished, if not altogether eliminated. Furthermore, a correct response does not involve verbalization; but at the same time ability to verbalize may be studied as an additional dependent variable.

The inductive method has been used in concept formation studies with children (Heidbreder, 1928; Levy & Cuddy, 1956; Long, 1940; Welch & Long, 1942, 1943) and adults (Archer, Bourne, & Brown, 1955), but in most of these studies *S* was restricted to a relatively small number of trials. Welch and Long (1943), for example, allowed only 30 trials for a problem involving Mills' joint method, and then concluded that no *S* was able to solve problems of this type. In the study by Levy and Cuddy, a rote learning component was introduced in a modification of the inductive method in which only 10 different stimuli were presented over and over again. In the investigation described below 150 trials were given (unless success was achieved earlier). A pilot study revealed that only one or two *S*s out of 100 who do not succeed within 150 trials attain success if given additional opportunity. A sufficiently large number of trials thus provides the possibility not only for the study of success in attaining the concept but also for estimating the relative difficulty of the concept as a function of other variables, such as age and intelligence.

In fact, one of the aims of the present investigation was to determine the role of intelligence in concept attainment by induction. While it seems reasonable to assume that intelligence and concept attainment are highly correlated, such a relationship has not been consistently demonstrated (Eisman, 1958; Griffith, Spitz, & Lippman, 1959; Hoffman, 1955; Plenderleith, 1956; Stuart, 1935). Furthermore, most studies have compared normal and feebleminded *S*s, while in the present investigation a comparison is made between normal and superior *S*s. Three age groups were selected for this study: 6-, 10-, and 14-yr. *S*s. These age groups were chosen because they represent the extremes and midpoint of the elementary school population and because they fall within three stages of mental development as defined by Piaget.

The concepts studied were: bird, animal, and living thing. According to the notions of Welch and Long (1940) these three concepts should represent increasing levels of difficulty, and thereby enhance the discriminating power of the tasks.

METHOD

Subjects.—The *S*s consisted of 180 school children attending elementary and junior high schools of Baltimore City and County, equally divided among the following age groups 6-, 10-, and 14-yr. Each *S* was within 6 mo. of the age specified. Half of the children within each age level had a WISC IQ between 90 and 109 (mean IQ for total normal intelligence group was 101.6, $SD = 5.1$) and the other half had a WISC IQ of 110 and above (mean IQ for total high intelligence group was 121.3, $SD = 8.1$).

Apparatus.—The apparatus consisted of two 3-in. hollow wooden cubes painted black. One side of the cube was missing so that it could be placed over a marble. A metal frame was attached to the side of the cube facing *S* in such a way that 2.75-in. cardboard stimuli could easily be slipped into it. A set of stimuli for each concept consisted of 150 pairs of pictures individually mounted on squares of white cardboard. One picture of each pair represented an example of the concept; the other picture was similar in area, color combination, and brightness, but represented an object not belonging to the concept category. A different set of stimulus pairs was used to represent each of the three concepts, bird, animal, and living thing, making a total of 900 stimulus cards. Figure 1 shows some of the stimulus pairs. It should be remembered, however, that the actual stimuli were multicolored.

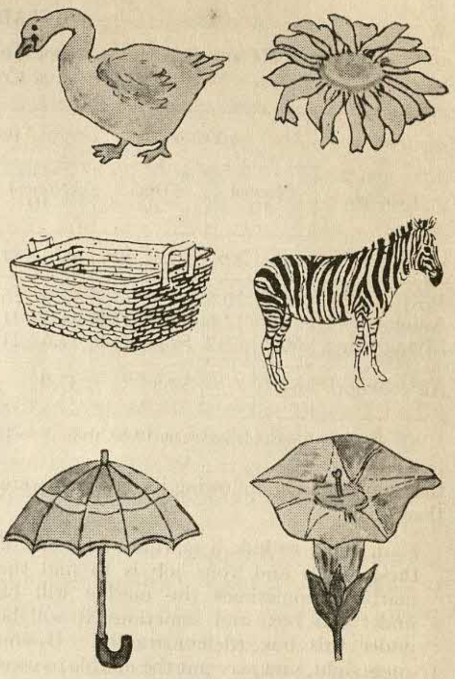


FIG. 1. Samples of stimulus pairs used in the bird, animal, and living thing concepts, respectively.

The reason for choosing naturalistic stimuli rather than formal types which would allow for systematic variation of dimensions was to avoid the complication of unequal familiarity or availability of verbal labels, which would tend to favor the older and more intelligent *S*s. Prior to the actual investigation the stimulus material was shown to 10 children in each age and intelligence group to be tested. These *S*s were asked to identify the pictures on the cards and any picture which was not properly identified was discarded. The children used in the pretest did not serve as *S*s in the experiment. The familiarity of the material was further checked by asking each *S* to identify 10 pairs of stimuli upon completion of the concept learning task. Different sets of 10 were used for each *S*. These data confirmed the fact that the pictures were entirely familiar to all *S*s.

A black curtain, which was drawn closed between responses, was attached to a rod mounted on a frame and placed on the table between *E* and *S*. The *S* was provided with a grooved wooden tray for collecting the marbles with which his correct responses were reinforced.

Procedure.—Each *S* was tested individually in two separate sessions. At the first session *S* was told that he would play a game in which he could win any one of a group of toys displayed before him. He was asked to select the most preferred of the group, and this was then set slightly apart from the others in a position which allowed him to view it during

TABLE 1

MEAN NUMBER OF ERRORS PER GROUP FOR REACHING CRITERION OF 10 CONSECUTIVE CORRECT RESPONSES

Concepts	6-Yr. *S*s				10-Yr. *S*s				14-Yr. *S*s				All Ages	
	Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ	High IQ
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*		
Bird	55.7	20.6	54.9	23.7	32.0	26.4	19.5	26.4	27.7	20.8	10.0	13.5	38.5	28.1
Animal	70.3	12.6	55.7	27.8	34.2	21.0	17.8	26.1	29.1	21.2	12.6	10.8	44.5	28.7
Living thing	66.2	5.9	50.9	23.0	47.6	21.2	38.7	22.0	30.8	28.2	9.4	10.5	48.2	33.0
All concepts	64.1		53.8		37.9		25.3		29.2		10.7		43.7	29.9

Note.—Each mean is based on 10 *S*s, total $N = 180$.

the "game." The following instructions were then given:

> I am going to hide a marble under one of these boxes and your job is to find the marble. Sometimes the marble will be under this box, and sometimes it will be under this box (demonstrated). If you guess right, you may put the marble in your tray. If you don't guess right, I will keep the marble. I will put cards in here like this (demonstrated with sample cards unrelated to the concept studied). If you look at the cards, you will be able to get many marbles. When you get enough marbles to fill your tray you will get the (name of the toy).

The *S* was required to point to the box which he thought covered a marble. The *E* then lifted the box, and if the response was correct, *S* picked up the marble and placed it in his tray. The procedure was continued until *S* made 10 consecutive correct responses, at which point *E* commented on *S*'s good performance and asked him how he knew the location of the marble.[3] The maximum number of trials was 150; and any *S* who did not reach the criterion of success at that point was considered to have failed. Each *S* worked on one concept only. The second session, held approximately two weeks later, was devoted to the administration of the WISC.

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS

Source	*df*	*MS*	*F*
Concepts (C)	2	801.58	1.620
Age (A)	2	24,053.90	48.616***
Intelligence (I)	1	8,556.00	17.293***
C × A	4	753.76	1.523
C × I	2	135.84	—
A × I	2	274.24	—
C × A × I	4	150.73	—
Within	162	494.77	
Total	179		

*** $P < .001$.

These experiments were conducted in schools; and one serious problem that arose was communication among the children. To prevent this the number of *S*s used in any one school was small; and *S*s, who had been chosen at random from the appropriate age groups, were tested in sequences which were likely to minimize communication. A split-half comparison of scores was made and no significant difference resulted.

Experimental design.—A 3 × 3 × 2 factorial design was used with the three age levels constituting one variable, the three concepts the second variable, and the two intelligence levels the third variable. This resulted in 18 groups with 10 *S*s in each group, consisting of approximately the same number of boys and girls.

RESULTS

The effect of age and intelligence.—The two response measures of errors to criterion and successful attainment of criterion were analyzed. Table 1 shows the mean number of errors per group and the corresponding *SD*s. These data did not satisfy the requirement of homogeneity for an analysis of variance. Nevertheless, an analysis of variance was performed for exploratory purposes; and the results are shown in Table 2. Even with the caution dictated by the lack of homogeneity, the *P* values associated with the effects of age and intelligence gave strong indication that these variables produced significant differences in performance.

---

[3] Originally it was planned to ask each *S* periodically during the experimental session for the basis of his choice. However, we discovered that frequently *S*, after stating his reason (such as "things with two legs" when working with birds), would change his strategy anticipating *E* not to bait stimuli that he expected *S* to choose.

The nonparametric Kruskal-Wallis one-way analysis of variance by ranks (Siegel, 1956) was also performed, and these results, shown in Table 3, were in complete agreement with the analysis of variance. To check out a possible interaction between age and intelligence an analysis was done on the data collapsed across concepts, thus producing homogeneity of variance. The *F* ratio for the Age × Intelligence interaction was less than one.

The distribution of *S*s in the 18 experimental groups who achieved criterion performance is shown in Table 4. An analysis of these frequency data, shown in Table 5, yielded results completely consistent with the error scores.

As an additional measure of association, the rank-order correlation was computed between error scores and mental age scores. As shown in Table 6, the rho values obtained are the same for each concept group. The data were, therefore, pooled and the value of the correlation for the total *S* population computed. All the values obtained were significant at the .001 level.

TABLE 3

KRUSKAL-WALLIS ANALYSIS OF VARIANCE OF ERROR SCORES

Source	*H*	*df*	*P*
Age	58.86	2	.001
Intelligence	14.83	1	.001
Concepts	4.27	2	*ns*

The effect of concept hierarchy.—From the results it appears that concept hierarchy had no effect on one measure of performance, errors to criterion, and only a minimal effect on frequency of *S*s who achieved criterion performance. As shown in Table 5, the concepts did produce a significant difference in the Normal Intelligence group. This analysis was done according to Cochran (1954) to take advantage of the predictions of the differences to be expected. The *S*s of high intelligence found all concepts of equal difficulty.

DISCUSSION

That age and intelligence are associated with more effective concept learning

TABLE 4

NUMBER OF *S*S PER GROUP WHO REACHED CRITERION PERFORMANCE

Concepts	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*		Total	
	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ
Bird	5	4	8	8	8	10	21	22
Animal	2	5	7	8	8	10	17	23
Living thing	0	4	5	8	7	10	12	22
Total	7	13	20	24	23	30	50	67

TABLE 5

Chi Square Analysis of Frequencies of Successful *S*s

Comparison	*df*	Chi Square
Age	2	42.636***
Intelligence	1	7.058**
Concept	1[a]	2.967
Concept (Normal IQ)	1[a]	5.468*
Concept (High IQ)	2	.116

[a] This calculation is based on Cochran's method, which involves loss of 1 *df* for the regression equation.
* $P < .02$.
** $P < .01$.
*** $P < .001$.

TABLE 6

Rank-Order Correlations between MA and Error Scores

Group	rho	*t*	*df*
Bird	−.58	5.41***	58
Animal	−.60	5.71***	58
Living thing	−.63	6.14***	58
Total	−.61	10.16***	178

*** $P < .001$.

seems entirely reasonable, but the specific mechanism by which the effectiveness is enhanced has yet to be clarified. Is the greater effectiveness of the older and more intelligent *S*s due merely to enhanced speed of learning, that is, to a greater increment in response tendency per trial, or is there also a qualitative difference in the learning process? To answer this question learning curves of individual *S*s were examined for the purpose of determining whether concept attainment was gradual or sudden. A gradual acquisition of the concept would be consistent with the continuity theory of learning and sudden acquisition with the theory of learning by hypothesis testing. In order to classify *S*s into gradual and sudden learners their performance just prior to the 10 consecutive correct responses was examined. Each *S* was assigned a score which represented the percentage of correct responses in the 10 trials preceding the final 10 (which were all correct). All *S*s whose score fell below the median, which was 60%, were classified as sudden learners, and those whose scores fell above the median were classified as gradual learners.[4] Table 7 shows the distribution of gradual and sudden learners, together with a chi square analysis, which shows that the incidence of sudden learners was a function of intelligence, but that neither age nor concept hierarchy produced any dif-

TABLE 7

Breakdown of Successful *S*s into Gradual and Sudden Learners

Concepts	6-Yr. *S*s				10-Yr. *S*s				14-Yr. *S*s				All Ages			
	Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ	
	S	G	S	G	S	G	S	G	S	G	S	G	S	G	S	G
Bird	1	4	3	1	3	5	6	2	2	6	7	3	6	15	16	6
Animal	2	0	2	3	4	3	7	1	4	4	7	3	10	7	16	7
Living thing	0	0	1	3	2	3	2	6	4	3	7	3	6	6	10	12
Total	3	4	6	7	9	11	15	9	10	13	21	9	22	28	42	25

Note.—S refers to sudden iearning, G to gradual learning. Age $\chi^2 = 1.066$, *df* = 2, not significant; Intelligence $\chi^2 = 4.034$, *df* = 1, $P < .05$; Concepts $\chi^2 = 2.731$, *df* = 2, not significant.

---

[4] Five *S*s who fell into the gradual category had each achieved the concept in 20 trials or less and made no more than two errors. These *S*s were arbitrarily placed into the sudden category, as their performance seemed to justify. This accounts for the unequal number of *S*s in the sudden gradual categories. Two of the *S*s had normal and three had high intelligence.

ference. From this finding an association between intelligence and concept attainment by hypothesis testing was inferred; an inference which is entirely consistent with recent findings by Kendler and Kendler (1959) and Kendler, Kendler, and Wells (1960).

One other factor that may be related to performance must be considered. The instructions given to *S*s did not direct them explicitly to try to discover a concept; but merely stated that attending to the stimuli would help them win marbles. Theoretically, therefore, it is possible that the experimental task consisted of two stages: (*a*) the discovery of the problem, that is, that a concept is to be sought, and (*b*) the solution of the problem, that is, the discovery of the specific concept. The differences in performance associated with age and intelligence may be due to more effective performance at either or both of these stages. To determine whether problem finding is a significant variable and the extent of its influence upon performance, if any, would require a systematic study of the effect of instructions upon concept attainment. Recent findings on the effects of stimulus complexity upon concept attainment at two levels of intelligence (Osler & Trautman, 1961) support the interpretation of a difference in the process of concept attainment as a function of intelligence.

A few brief remarks concerning the correlation between errors and mental age are in order. The rank-order correlation was selected because the data did not meet the assumptions for either the product moment or the correlation ratio. The actual value of rho obtained is probably lower than the true value because of the restricted range of talent in the present sample. The range of mental ages was restricted in terms of the ages studied and the minimum IQ required; the error scores were restricted because of the maximum number of trials permitted. The relatively high values of rho obtained, despite these limitations, emphasize further the importance of the variables studied.

SUMMARY

This investigation was concerned with the effect of age and intelligence on concept attainment. The *S*s were 180 children, 60 in each of the following age groups: 6, 10, and 14 yr. They were divided equally into normal and high intelligence groups. Each *S* worked on one of the following concepts: bird, animal, or living thing.

The response measures of trials to criterion and the number of successful *S*s were analyzed. The following conclusions were drawn: (*a*) Age was associated with significant differences in errors to criterion and number of successful *S*s. (*b*) Intelligence was associated with significant differences in errors to criterion and number of successful *S*s. (*c*) The effect of concept hierarchy was nominal. It had no effect upon errors to criterion; but it did produce significant differences in number of *S*s of normal intelligence who achieved the criterion of success. (*d*) When *S*s were divided into sudden and gradual learners, it was found that the frequency of sudden learners was a function of intelligence, but not of age. This was interpreted to indicate the greater utilization of mediators by the superior *S*s. (*e*) Rank-order correlations were computed between mental age and error scores and these were found to be highly significant.

REFERENCES

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

BRIAN, C. R., & GOODENOUGH, F. L. Relative potency of color and form perception at various ages. *J. exp. Psychol.*, 1929, **12**, 197–213.

COCHRAN, W. G. Some methods for strengthening the common χ^2 tests. *Biometrics*, 1954, **10**, 417–451.

COLBY, M. G., & ROBERTSON, J. B. Genetic studies in abstraction. *J. comp. Psychol.*, 1942, **33**, 385–401.

DATTMAN, P. E., & ISRAEL, H. E. The order of dominance among conceptual capacities: An experimental test of Heidbreder's hypothesis. *J. Psychol.*, 1951, **31**, 147–160.

EISMAN, B. S. Paired associate learning, generalization and retention as a function of intelligence. *Amer. J. ment. Defic.*, 1958, **63**, 481–489.

GRIFFITH, B. C., SPITZ, H. H., & LIPMAN, R. S. Verbal mediation and concept formation in retarded and normal subjects. *J. exp. Psychol.*, 1959, **58**, 247–255.

Hazlitt, V. Children's thinking. *Brit. J. Psychol.*, 1930, **20**, 354–361.

Heidbreder, E. Reasons used in solving problems. *J. exp. Psychol.*, 1927, **10**, 397–414.

Heidbreder, E. Problem solving in children and adults. *J. genet. Psychol.*, 1928, **35**, 522–545.

Heidbreder, E. The attainment of concepts: III. The process. *J. Psychol.*, 1947, **24**, 93–138.

Heidbreder, E. Experiments by Dattman and Israel on the attainment of concepts. *J. Psychol.*, 1952, **34**, 115–136.

Hoffman, H. N. A study in an aspect of concept formation, with subnormal, average, and superior adolescents. *Genet. psychol. Monogr.*, 1955, **52**, 191–239.

Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**(1, Whole No. 123).

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

Klingberg, G. The distinction between living and not living among 7–10 year old children, with some remarks concerning the so-called animism controversy. *J. genet. Psychol.*, 1957, **90**, 227–239.

Levy, N. M., & Cuddy, J. M. Concept learning in the educationally retarded child of normal intelligence. *J. consult. Psychol.*, 1956, **20**, 445–448.

Long, L. Conceptual relationships in children: The concept of roundness. *J. genet. Psychol.*, 1940, **57**, 289–315.

Long, L., & Welch, L. Reasoning ability in young children. *J. Psychol.*, 1941, **12**, 21–44.

Osler, S. F., & Trautman, G. E. Concept attainment: II. The effect of stimulus complexity on concept attainment at two levels of intelligence. *J. exp. Psychol.*, 1961, **62**, 9–13.

Piaget, J. *The child's conception of the world.* New York: Harcourt, Brace, 1929.

Piaget, J. *The child's conception of physical causality.* New York: Harcourt, Brace, 1930.

Plenderleith, M. Discrimination learning and discrimination reversal learning in normal and feebleminded children. *J. genet. Psychol.*, 1956, **88**, 107–112.

Russell, R. W. Studies in animism: II. The development of animism. *J. genet. Psychol.*, 1940, **56**, 353–366.

Russell, R. W. Studies in animism: V. Animism in older children. *J. genet. Psychol.*, 1942, **60**, 329–335.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Stevenson, H. W., & McBee, G. The learning of object and pattern discriminations by children. *J. comp. physiol. Psychol.*, 1958, **51**, 752–754.

Stuart, H. N. A study of sensorimotor and conceptual thinking in children between the ages of nine and eighteen. *J. exp. Educ.*, 1935, **4**, 147–153.

Thompson, J. The ability of children of different grade levels to generalize on sorting tests. *J. Psychol.*, 1941, **11**, 119–126.

Vinacke, W. E. *The psychology of thinking.* New York: McGraw-Hill, 1952.

Welch, L. The genetic development of the associational structures of abstract thinking. *J. genet. Psychol.*, 1940, **56**, 175–206.

Welch, L., & Long, L. The higher structural phases of concept formation of children. *J. Psychol.*, 1940, **9**, 59–95.

Welch, L., & Long, L. Methods used by children in solving inductive reasoning problems. *J. Psychol.*, 1942, **14**, 269–275.

Welch, L., & Long, L. Comparisons of the reasoning ability of two age groups. *J. genet. Psychol.*, 1943, **62**, 63–76.

(Early publication received
December 28, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 9–13

CONCEPT ATTAINMENT: II. EFFECT OF STIMULUS COMPLEXITY UPON CONCEPT ATTAINMENT AT TWO LEVELS OF INTELLIGENCE[1]

SONIA F. OSLER AND GRACE E. TRAUTMAN

Goucher College

Department of Pediatrics, Johns Hopkins University

Recent experimental work has lent support to the notion that discrimination learning and concept attainment by children may be achieved either by mediational symbolic processes or simple S-R associations. The experimental evidence also points to an association between the specific learning process characteristic of *S* and his intelligence. In a study of discrimination learning Kendler and Kendler (1959) found that fast learners achieved a reversal shift more readily than a nonreversal shift, while the opposite was true of slow learners. From this result it was inferred that fast learners utilized mediators in the original discrimination, while the slow learners achieved the discrimination through an S-R association.

In a second study involving concept attainment by induction (Osler & Fivel, 1961) *S*s were divided into gradual and sudden learners on the basis of the learning curves they generated. It was then found that in the group of superior intelligence there were more sudden learners than gradual learners, while the reverse was true in the group of normal intelligence. On the assumption that the gradual learners build up S-R associations and that sudden learners test successive hypotheses, it is concluded that the process mediating concept attainment is a function of intelligence.

The object of the present investigation was to provide an experimental means for testing the inferred relationship between intelligence and the specific learning mechanism involved in concept attainment. If hypothesis testing is more frequent among superior than normal *S*s, it should be possible to influence the performance of the superior group by varying the number of irrelevant dimensions, on which hypotheses can be based, in the concept exemplars. For *S*s of normal intelligence, who tend to achieve solution by the gradual building up of an S-R association, no systematic relation between the number of stimulus dimensions and speed of solution is anticipated.

The specific prediction, therefore, is that the performance measures in concept attainment will show a significant interaction between level of intelligence and the number of irrelevant dimensions in the concept exemplars.

---

[1] This investigation was supported by a grant from the National Institute of Mental Health, Psychopharmacology Service Center, made to Leon Eisenberg of Johns Hopkins University, and the senior author (Grant MY-2583).

The writers wish to express their gratitude to Mary A. Adams, Assistant Superintendent, Elementary Education, Lorne H. Woollatt, Former Director of Bureau of Research, Vernon S. Vavrina, Director, Junior and Senior High Schools, and to Arthur Lichtenstein, Director of Special Services for Pupils, Baltimore City Department of Education; and to B. Melvin Cole, Director, Elementary Education, Norris A. King, Director, Secondary Education, and Charles J. Leiman, Supervisor, Psychological Services, Baltimore County Department of Education; and to the principals and their staffs who cooperated in the conduct of the study.

They also wish to thank Nancy A. Freydberg, Sandra J. Weiss, and Frances Odette for testing *S*s and for many suggestions concerning the conduct of the study.

METHOD

Stimuli.—One concept was studied: the number *two* represented by two types of exemplars. The first method of representation consisted of two solid black circles, ¼ in. in diameter, placed in random fashion on a 5 × 5 grid on white cards 2¾ in. square. The complete set consisted of 150 different patterns; these were called the *formal two* (FT) stimuli. The negative stimuli in this set consisted of the same black circles in numbers varying from one to five, excluding two, placed in unduplicated random patterns.

The second method of representation was by means of two identical pictures of common objects pasted on cards identical to those used in the FT set. The pictures, which were in a variety of colors and sizes, were cut out from children's books and were different for each of the 150 cards. This set of stimuli was called the *object two* (OT). The negative stimuli for this set consisted of the same types of pictures in numbers other than two. An effort was made to have the negative stimuli match their positive mates in general area, brightness, and color combinations. Figure 1 contains examples of the two types of stimuli; but it should be remembered that the OT stimuli were actually multicolored. Altogether there were 600 different stimulus cards.

The reason for choosing naturalistic stimuli, rather than attempting to increase the complexity of the FT stimuli by introducing the variables of color or shape, was to create very large differences in the number of irrelevant dimensions in the concept exemplars without increasing their complexity to a point which gave the children of superior intelligence or older age an advantage in familiarity or in the availability of verbal labels. The use of common objects, pretested for familiarity, permitted the introduction of an exceedingly high number of attributes without the complications just enumerated. The weakness of this method of stimulus construction is that there is no control of the amount of irrelevant information; but for the purpose of this experiment this does not seem to be crucial, as long as it was reasonable to assume a *difference* in the amount of irrelevant information.

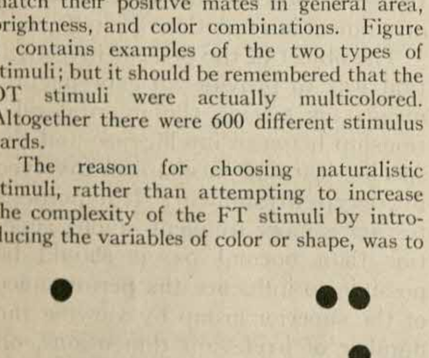


FIG. 1. Samples of the stimulus pairs used in the Formal Two (FT) and the Object Two (OT) concepts, respectively.

Prior to the actual investigation the stimulus material representing the object two was shown to 10 children in each age and intelligence group to be tested. These children were asked to identify the pictures and those which were not properly identified were discarded. The children used in this pretest did not serve as *S*s in the experiment. The familiarity of the material was checked further by asking each *S* to identify 10 pairs of stimuli upon completion of the concept learning task. Different sets of 10 were used for each *S*. These data confirmed the fact that the pictures were entirely familiar to all *S*s.

The rationale for the design of the stimuli was that the OT exemplars would suggest many more hypotheses than the FT exemplars because of the diversity in the shape, size, color, and content of the different objects represented. Preliminary work had shown that children gave three times as many different reasons for their choice of the OT as for the FT stimuli. For this reason it was anticipated that *S*s of superior intelligence would find the OT stimuli more difficult than the FT stimuli, while this should not hold for *S*s of normal intelligence.

Subjects.—The *S*s consisted of 120 school children attending elementary and junior high schools of Baltimore City and County, equally divided among 6, 10, and 14-yr. groups. Each *S* was within 6 months of the age specified. Half of the *S*s within each age level had a WISC IQ of 90 to 109 (the mean IQ of the total normal intelligence group was 101.3, $SD = 5.73$) and half had an IQ of 110 or above (the mean IQ of the total high intelligence group was 119.7, $SD = 7.46$).

Apparatus and procedure.—The apparatus and procedure have been completely described by Osler and Fivel (1961). Each *S*, who worked individually, was asked to choose the correct one of two stimuli. Reinforce-

TABLE 1

MEAN NUMBER OF ERRORS PER GROUP FOR REACHING CRITERION OF 10 CONSECUTIVE CORRECT RESPONSES

Stimuli	6-Yr. *S*s				10-Yr. *S*s				14-Yr. *S*s				All Ages	
	Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ		High IQ		Normal IQ	High IQ
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	Mean
Object Two	48.6	29.5	55.1	25.3	42.1	23.4	49.1	27.2	18.2	20.2	19.3	26.3	36.3	41.2
Formal Two	48.0	28.9	36.6	16.1	47.6	24.0	23.6	29.0	15.5	19.3	9.7	12.7	37.0	23.3

Note.—Each mean is based on 10 *S*s, total $N = 120$.

ment which followed the choice of the concept exemplar consisted of marbles to be exchanged at the end of the experiment for a prize of *S*'s selection. The *S* worked until he gave 10 consecutive correct responses, which was the criterion of success. The maximum number of trials was 150 and any *S* who did not achieve criterion performance at that point was considered to have failed. A second session, held within two weeks of the first, was devoted to the administration of the WISC.

Experimental design.—A 3 × 2 × 2 factorial design was used, with three ages constituting one variable, the two methods of concept representation the second variable, and the two intelligence levels the third variable. This resulted in 12 groups with 10 *S*s in each group, consisting of approximately the same number of boys and girls.

RESULTS AND DISCUSSION

The results were analyzed in terms of two measures of performance: (*a*) the number of errors per group, and (*b*) the number of *S*s who attained the criterion of success. Table 1 shows the mean number of errors for each of the 12 experimental groups with the corresponding *SD*s. An analysis of variance of these data, shown in Table 2, indicates that, of the three independent variables, only age produced significant differences; neither intelligence nor mode of concept representation as main effects produced significant *F* ratios. However, as predicted, the interaction between intelligence and method of concept representation turned out to be significant at the 5% level. A breakdown of the actual data shows that the interaction is due to the fact that *S*s of superior intelligence found the OT exemplars more difficult than the FT exemplars ($F = 7.739$, $df = 1/54$, $P < .01$); while *S*s of normal intelligence found both types of stimuli of equal difficulty. In fact, with the OT stimuli the superior *S*s lost all advantage of high intelligence, an advantage which was evident with the FT stimuli and also in the previous concept studies by Osler and Fivel (1961), utilizing a similar procedure.

The data on frequency of *S*s who achieved the criterion of success are shown in Table 3. A Chi square

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS

Source	*df*	*MS*	*F*
Age (A)	2	10,994.28	17.09**
Intelligence (I)	1	589.63	
Stimuli (S)	1	2,201.63	3.42
A × I	2	124.06	
A × S	2	44.31	
S × I	1	2,594.70	4.03*
A × S × I	2	363.93	
Error (Within)	108	643.45	
Total	119		

* $P < .05$.
** $P < .001$.

TABLE 3

NUMBER OF *S*S PER GROUP WHO REACHED CRITERION PERFORMANCE

Stimuli	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		All Ages	
	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ	Normal IQ	High IQ
Object Two	4	3	6	4	9	7	19	14
Formal Two	5	8	6	8	9	10	20	26

analysis shows that age ($\chi^2 = 15.60$, $df = 2, P < .001$) and concept representation ($\chi^2 = 6.26, df = 1, P < .02$) produced significant differences, while intelligence ($\chi^2 = .036$) did not. Because of the significant interaction in the error scores between intelligence and mode of concept representation, the frequency data were analyzed separately for each intelligence group. This analysis showed that for the normal *S*s concept representation made no difference ($\chi^2 = .074$), while for the superior *S*s the OT stimuli were more difficult than the FT stimuli ($\chi^2 = 10.80$, $df = 1$, $P < .01$). The results of the analysis of the frequency data are in complete agreement with those obtained from the error scores.

It has been demonstrated that the effect of increasing the number of irrelevant stimulus dimensions was to slow down *S*s of superior intelligence without affecting those of normal intelligence. These findings are consistent with the interpretation that in concept attainment high intelligence is associated with hypothesis testing, while normal intelligence is characterized by associative learning.

In the experiment by Kendler and Kendler (1959) it was found that the use of mediating responses speeded up learning. Is the present finding of no difference between normal and superior *S*s with the OT stimuli inconsistent with their finding? Any seeming inconsistency is easily dispelled by pointing out that where the number of irrelevant dimensions is small, as is the case with the FT stimuli, the present results are in complete accord with those reported by Kendler and Kendler, who also used simple stimuli. The more complex OT stimuli present a different problem, as is evident from the previous discussion.

Bourne and Restle (1959) recently suggested an inverse relationship between the number of irrelevant dimensions and the speed of concept identification, without restrictions as to *S* variables. They attempted a validation of this theoretical formulation with data from several studies, all of which used college students as *S*s. Since these *S*s are of superior intelligence, the findings are entirely consistent with those obtained in the present study for the high intelligence group.

But what about the group of normal intelligence for which the relationship proposed by Bourne and Restle did not hold? In casting about for a possible explanation one might ask whether the confounding of complexity with thematic quality of the stimuli used in the present investigation might not have altered the experimental situation. Theoretically it may have; but it is not immediately apparent why the effect of thematic stimuli should have been different for the two intelligence groups. Bruner, Goodnow, and Austin (1956, p. 108), who investigated selection strategies in concept attainment using abstract and thematic material, report that thematic material retarded the performance of their *S*s by generating more incorrect hypotheses. Their *S*s, who were Harvard undergraduates, are comparable to the superior group of the present study. Again, their results remain consistent with those reported here; but still this throws no additional light on the findings with the normals. It may be that either Bourne and Restle should not extrapolate from their data obtained with superior *S*s to those of normal intelligence; or that there is an interaction effect with the age variable whereby additional dimensions produce effects upon adults but not upon children of normal intelligence.

SUMMARY

The aim of this investigation was to provide a means of testing a previous interpretation that *S*s of superior intelligence attain concepts by testing hypotheses while *S*s of normal intelligence attain concepts through S-R associative learning. One hundred and twenty *S*s, 60 at each intelligence level, were tested on a concept attainment task. Half the *S*s worked with simple stimuli and half with complex stimuli. In each case the concept to be attained was the same, that of the number two. Because the complex stimuli would generate more hypotheses than the simple ones, it was predicted that they would slow down the superior *S*s, but not affect the normal *S*s. On two measures of performance, number of errors to criterion and frequency of *S*s attaining criterion, the prediction was confirmed.

REFERENCES

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, 66, 278–296.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58,** 56–60.

OSLER, S. F., & FIVEL, M. W. Concept attainment: I. Effect of age and intelligence in concept attainment by induction. *J. exp. Psychol.*, 1961, **62,** 1–8.

(Early publication received
December 28, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 14–23

FAMILIARIZATION (n) AS A STIMULUS FACTOR IN PAIRED-ASSOCIATE VERBAL LEARNING[1]

DONALD R. GANNON AND CLYDE E. NOBLE

Louisiana State University *Montana State University*

The Law of Acquaintance is currently enjoying a scientific revival among human learning psychologists. According to Robinson: "the act of reading or reciting . . . may . . . influence the facility with which that act enters into new associations" (1932, p. 118). Despite its premature announcement nearly thirty years ago, very little evidence favoring this principle was available until the past decade when the quantitative effects of frequency of stimulation (n) upon recognition behavior and familiarity judgments were discovered (Arnoult, 1956; Noble, 1954, 1960; Solomon & Postman, 1952; Underwood & Schulz, 1960, Exp. I). In general, the facilitation functions for a variety of materials are negatively accelerated in the range $n=0$ to 40, with returns diminishing rapidly around $n = 20$. Serial learning proficiency (R%) is similarly affected by n (Hovland & Kurtz, 1952; Noble, 1955; Riley & Phillips, 1959; Underwood & Schulz, 1960, Exp. III), but paired-associate learning presents a maze of contradictions.

One of G. E. Müller's students (Winzen, 1921) made an early attempt to determine whether familiarization is more influential in the stimulus (S) or in the response (R) position. Learning pronounceable syllables which they had seen either 0 or 20 times the day before, *S*s practiced by the method of right associates (Treffermethode) at a presentation rate of about one S-R pair every 9 sec. Winzen reported that the combination S_{20}-R_0 produced more correct responses than the combination S_0-R_{20}. Unfortunately, the control pairs S_0-R_0 and S_{20}-R_{20} were not included in the experimental design, nor was the statistical treatment adequate by modern standards to permit any definite conclusions. Two recent studies (Cieutat, 1960; Morikawa, 1959, Exp. II) lend partial support to Winzen's observations, but other investigators have either found null results (Bailey & Jeffrey, 1958; Mandler & Campbell, 1957; Waters, 1939) or contradictory evidence (Sheffield, 1946; Underwood & Schulz, 1960; Weiss, 1958). These experiments will be discussed later.

The present study was designed to evaluate an hypothesis based on data pertaining to *meaningfulness* (Noble, 1952a) available in 1957. Two independent experiments by Cieutat, Stockwell, and Noble (1958), confirmed since by others (Cieutat, 1959; Hunt, 1959; Morikawa, 1959, Exp. I), showed significant effects of meaningfulness on both the S term and the R term but a greater influence on R than on S. Mindful of the serial learning data, they reasoned that if differences in the learning of paired associates varying in meaningfulness are due primarily to the positive transfer effects of *n*, an appropriate experiment

---

[1] This paper is based upon a master's thesis performed by DRG under the guidance of CEN. An abridged report was made to the Southeastern Psychological Association, April 1958. We are grateful to Blaine L. Baker, Victor J. Cieutat, and Charles H. Koski for checking the statistical analyses; to Irmgard Taylor for assistance in translation of the Winzen article; and to Margaret B. Koch, Donald J. Lewis, and Ruth H. Preston for helpful suggestions.

Preparation of this report was facilitated by a contract between the Office of Naval Research and Montana State University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

> might reveal that meaningfulness is a secondary phenomenon derivable from the same basic variable that controls all learning. This coordination would be demonstrated neatly if one could produce similar acquisition curves by manipulating either (frequency or meaningfulness). The empirical question [they concluded] is whether familiarization would affect the S and R terms differentially (1958, p. 201).

Since it is known that R% is a function of meaningfulness in serial and paired-associate learning (Noble, 1952b; Noble & McNeely, 1957); that R% is a function of familiarization in serial learning; and that the meaningfulness of the R term in paired associates contributes more to learning proficiency than the S term, it follows that if familiarization is the *sole* basis of the meaningfulness effect, then familiarization should facilitate the learning of paired associates in a similar manner. In symbolic terms, the following three implications were drawn by analogy to the two experiments by Cieutat, Stockwell, and Noble (1958): (*a*) $(S_{20}\text{-}R_{20}) > (S_0\text{-}R_{20}) > (S_{20}\text{-}R_0) > (S_0\text{-}R_0)$. (*b*) All main effects due to the stimulus term (S), the response term (R), and number of practice trials (N) should be significant. (*c*) Given enough practice, there ought to be significant two-factor interactions (S × R, S × N, R × N) and a three-factor interaction (S × R × N).

Should the main effects and interactions turn out radically different from the predictions above, it would incline one toward the alternative hypothesis that familiarization (*n*) cannot be used to explain all of the variance in R% scores thought to be due to meaningfulness (*m*), and that there must be more to meaningfulness than sheer frequency of experience, e.g., number of associations (Noble, 1961; Parker & Noble, 1960). The latter is the view of *m* originally espoused by the second author (Noble, 1952a); hence the present investigation is at once an indirect test of the associationistic concept of meaningfulness and a direct test of Winzen's stimulus-familiarity hypothesis.

METHOD

Design considerations.—The experiment consisted of two phases: familiarization and learning. In the familiarization phase *S*s received either 0 or 20 independent exposures (*n*) of items of low meaningfulness (*m*) and familiarity (*f*). The "words" later to be designated as S and R terms were never paired; each was presented as a single stimulus to which *S* made the identifying response of pronunciation. This insured stimulus reception as well as exercise of the response. The learning phase was designed to determine whether *n* facilitated learning, and if so, whether familiarization differentially influenced the relative contributions of S and R to the learning of five pairs of the previously isolated items. Four experimental groups were formed by all combinations of locus (S vs. R) and amount (0 vs. 20) of familiarization, while a control group received no familiarization training. All *S*s practiced the same five S-R pairs during the learning phase, so any observed differences in R% scores among the experimental groups would have to be attributed to effects of the familiarization phase.

Several methodological points must be considered. First, high *m* theoretically requires high *n*, but not conversely (Noble, 1953); familiarization effects in this study may not, therefore, be explained as changes in meaningfulness. One experiment consistent with this view (Riley & Phillips, 1959) found that although serial learning of CVC material was accelerated by prior familiarization, the *m* values of the CVCs were unaffected. Second, it is important in the present investigation to use independent groups of naive *S*s in a constant-trials design. The purpose of this is to study acquisition phenomena including interactions, to avoid interference effects, to permit unbiased tests of S-R interaction, and to provide controls for generalized familiarization influences. Finally, in the interest of standardization, the familiarization and transfer operations are similar to those employed in earlier studies in this series.

Apparatus.—Instrumentation consisted of a Viewlex Viewmatic automatic slide projector (No. RCP-1) for the familiarization phase and a Patterson memory drum (No. 1-A) for the learning phase. The projector was equipped with a circular slide cartridge having a capacity of 30 2 × 2 in. slides. Four cartridges were required, each of which contained the familiarization items for a single experimental group. The memory drum was fitted with a white tape containing the paired associates typed in capital letters. A sliding shutter over the aperture permitted *S* to view either the left or the right side of the drum. The projection and tachistoscopic functions were controlled by two Hunter electronic timers set to expose each slide for 1 sec., and to provide a 1.4-sec. interstimulus period.[2] This gave a presentation rate of one slide every 2.4 sec. in the familiarization phase. For the learning phase, the intertrial interval was 8 sec., and the S item appeared for 2 sec. followed by the S and R items together for 2 sec.

TABLE 1

FAMILIARIZATION LISTS FOR THE FOUR EXPERIMENTAL GROUPS APPEARING IN EACH PROJECTOR CARTRIDGE

	S_{20}-R_{20}		S_0-R_{20}		S_{20}-R_0		S_0-R_0
I	DELPIN	I	DELPIN	I	DELPIN	I	DELPIN
	BYSSUS		BYSSUS		BYSSUS		BYSSUS
R	ULNA		ZUMAP	R	ULNA		JETSAM
	TAROP		JETSAM		TAROP		ZUMAP
	GOKEM		DELPIN		GOKEM		BYSSUS
	NOSTAW		QUIPSON		NOSTAW		QUIPSON
	SAGROLE		BYSSUS		SAGROLE		ZUMAP
	NOSTAW		ZUMAP		NOSTAW		DELPIN
	GOKEM		DELPIN		GOKEM		QUIPSON
	TAROP		JETSAM		TAROP		JETSAM
	ULNA		QUIPSON		ULNA		BYSSUS
	SAGROLE		BYSSUS		SAGROLE		DELPIN
I	JETSAM		JETSAM	I	JETSAM		JETSAM
	ZUMAP		ZUMAP		ZUMAP		ZUMAP
	QUIPSON		QUIPSON		QUIPSON		QUIPSON
	BYSSUS		BYSSUS		BYSSUS		BYSSUS
	ZUMAP		ZUMAP		ZUMAP		ZUMAP
	DELPIN		DELPIN		DELPIN		DELPIN
R	VOLVAP	R	VOLVAP		ZUMAP		QUIPSON
	XYLEM		XYLEM		JETSAM		DELPIN
	MEARDON		MEARDON		DELPIN		ZUMAP
	KUPOD		KUPOD		QUIPSON		JETSAM
	LATUK		LATUK		BYSSUS		DELPIN
	KUPOD		KUPOD		ZUMAP		BYSSUS
	MEARDON		MEARDON		DELPIN		QUIPSON
	XYLEM		XYLEM		JETSAM		JETSAM
	VOLVAP		VOLVAP		QUIPSON		BYSSUS
	LATUK		LATUK		BYSSUS		ZUMAP
I	QUIPSON	I	QUIPSON		QUIPSON		QUIPSON
	JETSAM		JETSAM		JETSAM		JETSAM
I	5@2 ea.		5@4 ea.		5@4 ea.		5@6 ea.
R	10@2 ea.		5@2 ea.		5@2 ea.		none
Σ	30		30		30		30

Note.—R = relevant items, I = irrelevant items.

Learning materials.—Fifteen dissyllabic "words," mostly paralogs, were used for familiarization training. Ten of these (relevant items) also appeared as the S-R elements of the paired-associate lists in the learning phase, while the remaining five (irrelevant items) were used only as fillers in the familiarization phase in order to equate the total number of exposures for *S*s in the four experimental conditions. All items were chosen from the low end of the *m* and *f* scales (Noble, 1953) so that the mean median *m* value was .30, and the mean *f* value was .41. Low scale values were used in order to reduce the number of possible S-R connections formed outside the experimental situation and to insure comparability with the previous serial learning study (Noble, 1955). The four familiarization lists are reproduced in Table 1.

Subjects.—The *S*s were 120 male and female college students obtained on a voluntary basis at Louisiana State University in 1957. Ages ranged from 17 to 43 yr. (mean = 19.3 yr.), and all *S*s were naive with respect to paired-associate learning. Twenty-four *S*s were assigned to each of the four experimental groups in a counterbalanced order of arrival at the laboratory. An additional 24 *S*s were used as a special control group which received no familiarization training. The *E* engaged these *S*s in conversation for an amount of time equivalent to the familiarization phase, then proceeded to the learning phase.

Procedure.—Four experimental groups were formed in terms of locus and amount of familiarization, as follows: S_{20}-R_{20}, S_0-R_{20}, S_{20}-R_0, S_0-R_0. In the familiarization phase, *S* was instructed to pronounce each dissyllable aloud as it was projected on the screen. The *S*s were not informed of the subsequent learning phase. As shown above in Table 1, the relevant items were separated by irrelevant items in order to reduce the formation of any initial S-R connections before the learning phase. The ratios of relevant to irrelevant items in the S_{20}-R_{20}, S_0-R_{20}, S_{20}-R_0, and S_0-R_0 groups, respectively, were 20/10, 10/20, 10/20, and 0/30. Since each relevant item was exposed twice in each cartridge, 10 revolutions were required to obtain 20 exposures. There were 300 exposures in all.

---

[2] The present exposure duration is from 1.3 to 2.0 sec. briefer than those used in earlier studies (Noble, 1954, 1955), but the interstimulus period is the same.

In the learning phase, *S* was instructed to anticipate the R term after the first trial. As is standard in this series, both S and R were pronounced and the correction method was used. Seventeen learning trials were given, with the two groups of five relevant items constituting the S and R terms. Responses were scored as correct (R+) or erroneous (R−). The criteria for an erroneous response were as follows: (*a*) omissions, (*b*) late responses, (*c*) mispronounced responses, (*d*) the same response repeated three times in succession during a single trial. The *S*s had to maintain a uniform pronunciation of the paralogs in order to receive R+ scores by the third error criterion. To neutralize accidental order effects, two lists of S-R pairs were formed. The Forward list, mounted on the left side of the memory drum, consisted of pairs formed from the 10 relevant items. The Reverse list, mounted on the right side of the drum, transposed these items so that S_1 became R_1, S_2 became R_2, etc. The *E* constructed four random trial sequences to minimize serial learning effects. These learning lists are reproduced in Table 2. There were three Forward and three Reverse *S*s assigned to each of the four trial sequences, making a total of 24 *S*s per condition. The familiarization and learning phases required 12 min. each, separated by a 1-min. rest interval during which *E* engaged *S* in conversation. Total time, including instructions, was approximately 27 min. for all *S*s.

Instructions.—The familiarization instructions were given verbatim as follows:

> This is an experiment on verbal learning. We are interested in the general learning process common to all people and are not testing your intelligence or personality. In the first part of the experiment you will be shown slides to get you acquainted with the procedure. With this projector, I am going to flash on the screen some two-syllable words similar to actual words. Probably you have never seen any of them before so there is no standard or correct pronunciation. All you have to do is say each word out loud before it leaves the screen. Whichever way you pronounce each word when we start is all right, but try to say it the same way each time the word comes up. Do *not* try to memorize the order of the words because this will change constantly. Please say each word loudly and clearly enough so I can hear and understand you. In this way I will be sure to give you credit for a response. Try to say the word before it leaves the screen, since the words come in rapid order. Are there any questions? Now remember, if you see "dinner," say "dinner." Ready?

TABLE 2

PAIRED-ASSOCIATE LISTS AND TRIAL SEQUENCES FOR ALL GROUPS DURING THE LEARNING PHASE

Trial Sequences	Learning Lists			
	Forward		Reverse	
I	TAROP SAGROLE GOKEM ULNA NOSTAW	XYLEM LATUK MEARDON VOLVAP KUPOD	XYLEM LATUK MEARDON VOLVAP KUPOD	TAROP SAGROLE GOKEM ULNA NOSTAW
II	ULNA GOKEM TAROP NOSTAW SAGROLE	VOLVAP MEARDON XYLEM KUPOD LATUK	VOLVAP MEARDON XYLEM KUPOD LATUK	ULNA GOKEM TAROP NOSTAW SAGROLE
III	GOKEM NOSTAW ULNA SAGROLE TAROP	MEARDON KUPOD VOLVAP LATUK XYLEM	MEARDON KUPOD VOLVAP LATUK XYLEM	GOKEM NOSTAW ULNA SAGROLE TAROP
IV	NOSTAW ULNA SAGROLE TAROP GOKEM	KUPOD VOLVAP LATUK XYLEM MEARDON	KUPOD VOLVAP LATUK XYLEM MEARDON	NOSTAW ULNA SAGROLE TAROP GOKEM

The learning instructions were given verbatim as follows:

> This is the second part of the experiment. Here is a memory drum. Shortly after the apparatus starts, you will see a word in the window. You are to pronounce this word and those that follow it as you see them. First, a word will appear alone, then the shade will rise and you will see another word. You are to pronounce them both. After you have seen this entire list once, your job will be to call out each single word and then to anticipate the word with which it was paired. In other words, as you see the *single* word, you are to pronounce it, and then immediately you are to pronounce the word with which it was paired *before* the second word appears. If you fail to anticipate an item, pronounce it when it appears anyway. If you think you know what a word will be but are not sure, *guess*. It will not hurt your score any more than to say nothing, and if you get it correct it will count as a success. If you anticipate a word *incorrectly*, correct yourself as soon as the word appears. Remember to pronounce each item *aloud*. Do not try to think ahead more than one step at a time or to count, because the *order* of the pairs will change as the list is repeated. Do not try to use any special system in your learning. Simply associate each single word with the word with which it is paired. Please try to pronounce all words the same way each time they appear so that I can give you credit for a correct response. Any questions?

RESULTS

The major results of the experiment are shown in Fig. 1. Percentages of correct responses (R%) are plotted as a function of amount of practice (N) grouped in blocks of two trials for the five conditions. The influence of S-term familiarization is greater

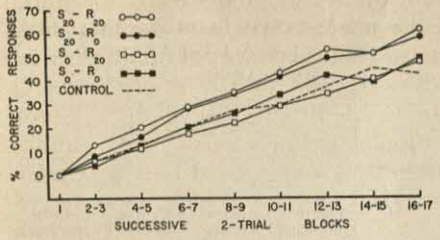


FIG. 1. Acquisition curves for a list of five paired associates as a function of practice under four different combinations of low ($n = 0$) and high ($n = 20$) frequency of stimulus (S) and response (R) term familiarization. (The Control group received no prior familiarization experience. Each curve contains 24 *Ss*.)

than that of R-term familiarization, while the effect of the latter treatment is little different from having no familiarization at all. Total R+ scores for the various groups during the 17-trial period, given in parentheses, were ordered as follows: S_{20}-R_{20} (735), S_{20}-R_0 (689), S_0-R_0 (542), Control (524), S_0-R_{20} (504). The S-term difference is (1424 − 1046 =) 378 units, contrasted with an R-difference of only (1239 − 1231 =) 8 units. An 8 × 5 Type I mixed-factorial analysis of variance (Lindquist, 1953, p. 267) was performed on the R+ scores from which Fig. 1 was constructed. There were significant main effects of Practice ($F = 157.29$; $df = 7/805$; $P < .001$) and Conditions ($F = 2.69$; $df = 4/115$; $P < .05$), but the interaction was not significant ($F < 1.00$; $df = 28/805$).

In order to test for a learning-to-learn familiarization effect, the Control group may be considered a baseline from which to measure possible generalized influences of the familiarization phase on the learning phase. A 3 × 1 simple-randomized analysis of variance (Lindquist, 1953, p. 47) was therefore applied to the overall R+ scores of the Control group and the two S_0 groups, omitting both S_{20} groups. The result gave no evidence of any nonspecific familiarization effect ($F < 1.00$; $df = 2/68$; $P > .20$), so the three lowest curves in Fig. 1 may be regarded as the same except for sampling fluctuations. There is neither a relevant response effect (S_0-R_{20} vs. S_0-R_0) nor an irrelevant stimulus-response effect (S_0-R_0 vs. Control) that can be attributed to the 300 trials of stimulation and pronunciation experienced during the familiarization phase.

To evaluate possible two-factor and three-factor interactions, a 16 × 2 × 2, N × S × R Type III mixed-factorial analysis of variance (Lindquist, 1953, p. 281) was performed on the individual trial R+ scores for the four experimental groups only. The summary, shown in Table 3, indicates significant main effects of Practice (N) and Stimulus Familiarization (S). Neither the Response Familiarization (R) factor nor any of the interactions were significant sources of variance.

Because R+ scores do not reveal any Practice × Conditions interactions, the facilitation due to S-term familiarization must be largely an initial effect followed by latter paral-

TABLE 3

ANALYSIS OF VARIANCE OF CORRECT RESPONSES (R+) OF THE EXPERIMENTAL GROUPS DURING TRIALS 2–17

Source	*df*	*MS*	*F*
Between *Ss*	95		
Stimulus (S)	1	93.02	7.86*
Response (R)	1	.04	.00
S × R	1	4.60	.39
Error (*b*)	92	11.83	
Within *Ss*	1440		
Trials (N)	15	57.99	84.30**
S × N	15	.70	1.03
R × N	15	.44	.63
S × R × N	15	.54	.78
Error (*w*)	1380	.69	
Total	1535		

* $P < .01$.
** $P < .001$.

lelism. This hypothesis is confirmed by comparing the average number of trials or errors (R−) preceding the first correct anticipation of each S-R pair in all groups. Most of these R− scores were omissions rather than late or overt incorrect responses, so no qualitative analysis of errors will be made. Mean R− scores required to attain this reaction-threshold criterion (averaged over *S*s not S-R pairs) were ordered as follows: S_{20}-R_{20} (5.20), S_{20}-R_0 (5.33), Control (6.28), S_0-R_{20} (6.88), S_0-R_0 (6.95). A 5 × 1 simple-randomized analysis of variance (Lindquist, 1953, p. 47) of these data gave an F ratio of 2.79 ($df = 4/112$; $P < .05$) which is significant. Again, the major difference favors the two S_{20} groups over the S_0 groups, with the Control group in the middle. Employing t tests of the deviations of each pair of extreme groups from the Control group based on the within-groups σ^2 (Lindquist, 1953, p. 91), we find significantly more errors under the S_0 (irrelevant *n*) combination and fewer errors under the S_{20} (relevant *n*) combination ($t = 3.31$; $df = 112$; $P < .01$); however, the two component ratios are not significant individually ($t \ngtr 1.69$; $df = 112$; $P > .05$). By the second correct (reinforced) anticipation all five groups have converged ($F < 1.00$; $df = 4/110$), and the speed of error elimination is essentially constant thereafter in agreement with Fig. 1. Observed group trends in R+ scores thus probably reflect unequal levels of proficiency (number of associates attained per trial) rather than differential rates of learning (growth of habit strength).

The results lead us to state that *n* has a facilitative influence on paired-associate learning only when familiarized items become the S members. Prelearning familiarization with the R members fails to produce any significant result, and except for initial error effects there is no further interaction between locus of familiarization (S vs. R) and amount of practice (N). In the light of these data, a simple stimulus-frequency theory of meaningfulness (Cieutat et al., 1958; Underwood & Schulz, 1960) must be rejected as untenable. We conclude that familiarization (*n*) does not produce effects which are equivalent to those produced by meaningfulness (*m*), and that *m* and *n* are therefore independent variables affecting verbal behavior.

DISCUSSION

Two principal questions are raised by this experiment: (*a*) Are the results comparable with those obtained by other investigators? (*b*) Assuming no artifacts, what is the most reasonable interpretation of the current data?

Comparability.—Eight laboratories have followed up Winzen's (1921) work. Negative results in the Waters (1939), Mandler and Campbell (1957), and Bailey and Jeffrey (1958) experiments seem to be contingent on the use of CVC or CCC materials and very brief amounts or durations of familiarization. Contradictory findings reported in the Sheffield (1946), Weiss (1958), and Underwood and Schulz (1960) experiments are also associated with trigram material; in addition, the types and amounts of familiarization as well as the transfer tests are different. In opposition to Winzen, the hypothesis coming from this second group of studies suggests that familiarization is effective only in the R position. Occasionally there were signs of decremental influences of S-term familiarization (Sheffield, 1946; Underwood & Schulz, 1960, Exp. I, IV), but these trends were neither significant nor consistent (Underwood & Schulz, 1960, Exp. II). Although the latter experiments contain many ingenious features, there are difficulties in evaluating the R-facilitation and S-interference phenomena because of rather complex familiarization techniques, sophisticated *S*s, and incomplete factorial designs. Statistical tests of interaction among the S, R, and N factors are frequently lacking. As Underwood and Schulz say of their Exp. I, II, and IV: "they have raised more issues than they have settled regarding the effects of stimulus familiarization" (1960, pp. 125–126).

In support of Winzen, on the other hand, are the present data together with those of Morikawa (1959, Exp. II) and Cieutat (1960). Morikawa noted definite S-facilitation in Japanese *S*s when the R term was unfamiliar but little difference when the R term was familiar. A different procedural approach was tried by Cieutat in that pronounceable items of low *m* value were used, but he varied exposure duration (0 vs. 60 sec.) of the two terms (S vs. R) rather than frequency and did not require verbalization. Another difference is that his *S*s served in all four combinations of duration and locus; however, both S and R were pronounced in the transfer period, as in this experiment, and the *S*s were from the same population. Cieutat found significant effects attributable to N and to the S × R interaction, the latter being observed as a weak S-term decrement under R_0, a strong S-term increment under R_{60}, and an overall main effect favoring stimulus familiarization. (Referring to his Table 2, mean R+ scores for the S_0 and S_{60} conditions were 2.31 and 2.50, respectively, whereas means for R_0 and R_{60} were both exactly 2.41 units.) Sustained viewing, possible intertreatment interferences, and failure to require articulation during the familiarization period may account for the slight (statistically marginal) inhibitory S-effect under irrelevant R-term familiarization; the absence of a reference level provided by an unfamiliarized control group precludes firm inferences. This peculiarity is also characteristic of the Sheffield and Underwood-Schulz experiments. In one crucial respect, however, Cieutat's data are quite consistent with Winzen's and ours; namely, the significant superiority of his S_{60}-R_0 treatment combination over that of S_0-R_{60}.

The use of confounded and incomplete designs, unrelated or inadequate control procedures, *S*s of heterogeneous experience, diverse types or amounts of familiarization, different learning materials, and varied methods of testing for transfer renders impossible any comparative evaluation of the present findings in the same context with the above studies. About all that can be done from a constructive point of view is to urge that the reader be cautious in trying to resolve these conflicting reports, and to judge each set of experimental data in the light of its own methodology. With the reminder that our procedure was standard in this research program (e.g., specific familiarization technique), and that it was designed to meet certain boundary conditions (e.g., two-term articulation during learning), we turn now to a tentative explanation of the results.

Interpretation.—The data permit three unequivocal statements: (*a*) relevant S-term familiarization produced significant superiority in total R+ scores whereas relevant R-term familiarization had no influence; (*b*) when compared with the unfamiliarized Control condition, initial R− scores increased after irrelevant S-term familiarization and decreased after relevant S-term familiarization; (*c*) acquisition rates of correct responses and extinction rates of incorrect responses both became independent of original familiarization following the first reinforcement in the list.

Recalling the introductory discussion, a two-factor hypothesis may be suggested to account for the differential effects of *m* and *n* in paired-associate learning under these conditions. The evidence indicates that *n* is sufficient to influence behavior via the S term alone (stimulus predifferentiation) whereas *m* is *jointly* necessary to affect behavior via the R term (response patterning). One possible explanation of S-term facilitation, considered earlier by the second author in connection with the complex familiarization effects in serial learning (where the "perceptual" and "motor" aspects of verbal learning are confounded), is provided by the principle of stimulus constancy (Noble, 1955, p. 336). The core assumption, derived from the conditioning literature, is that repeated exposure-articulation sequences during the familiarization phase will attenuate variability in *S*'s perceptual (e.g., orientation, fixation) and identifying (e.g., pronunciation) responses to the paired-associate S terms when the learning phase is encountered. Assuming further that the stability of these perceptual and identifying responses grows in some negatively accelerated fashion with increasing *n* (like habit strength or familiarity), their facilitative influences ought to be maximal early in the learning phase then subside later as paired-associate training progresses. Although less parsimonious than a simple frequency explanation, this notion helps to understand the consistent (but weak) S-effects of *m* which have been observed in a number of studies (Cieutat, 1959; Cieutat et al., 1958, Exp. I, II; Hunt, 1959; Morikawa, 1959, Exp. II; Underwood & Schulz, 1960).[3]

Our two-factor hypothesis implies a systematic view of the verbal learning process which is somewhat at variance with that held by Underwood and Schulz (1960). Their two-stage analysis of "response-recall" followed by "associative hook-up" (reminiscent of Thorndikean selecting and connecting) postulates that frequency alone (e.g., *n* without reinforcement) is important for the first stage of learning, the second being related to meaningfulness (or pronunciability). So far we might agree, but they proceed from their "spew" hypothesis, which is based on free-recall phenomena, yet considered "not relevant for stimulus familiarization" (p. 100), to predict that (*a*) "rate of paired-associate learning should be an increasing function of amount of familiarization training given the response units of the list," and (*b*) that "response familiarization should facilitate learning considerably more than stimulus familiarization" (pp. 100–101). The emphatic lack of any R-term effect or of an R > S relation in the present data is clearly embarrassing to the Underwood-Schulz hypothesis.

But is our evidence relevant? A sophisticated critic might object that this investigation (using pronounceable words and paralogs) is not a fair test of the "spew" hypothesis because it fails to satisfy some of the Underwood-Schulz boundary conditions (e.g., CVCs and the spelling method). Unfortunately, these writers are neither very specific nor consistent about such matters (cf. the familiarization procedures, transfer tests, and *S*s in Exp. I vs. Exp. IV), but in one chapter (p. 95) they describe a hypothetical task where, even though the responses are "well integrated" (pronounceable), prior differences in *n* are definitely expected to alter the "availability" (reaction threshold?) of the R terms. The present procedure appears to conform to these general requirements; only the data are recalcitrant, paradoxically. Considering their repeated emphasis on the first stage process, incidentally, we find it curious that Underwood and Schulz are less interested in initial performance and rate measurements than in overall analyses.

A second possible objection to this experiment turns on our custom in this series of having *S*s articulate both terms in the paired-associate learning phase. It may be claimed that this requirement either (*a*) produces an unusual amount of S-term familiarization, or (*b*) shortens the effective anticipation interval for the S_0 groups as contrasted with the S_{20} groups. Briefly (aside from the standardization issue), our rejoinder to (*a*) is that we believe double articulation is necessary both to guarantee stimulus reception and to provide an unbiased test of differences between the manipulated properties of the antecedent (S) and consequent (R) events. Since all factors except temporal succession should be equated when evaluating the effects

---

[3] This tentative explanation does not presume to account for the initial decrements in the S_0-R_0 and S_0-R_{20} groups (relative to the Control), but it would not be inconsistent with a theory of stimulus variability to hypothesize that irrelevant familiarization in the S-position produces interference via stimulus trace generalization.

EDNL. & PSY. RE
LIBRARY
1·9·70
9255

of *S*'s built-in response to an S term upon his ability to anticipate its paired R term, it seems elementary to have the former occur just as frequently as the latter during the learning phase (cf. also the rationale of the correction method on erroneous trials). This Underwood and Schulz did not do, hence it constitutes yet another difference between our procedures. Such a criticism could be turned around and directed at those who introduce biases favoring R-term familiarization.

The distributed-practice argument (*b*) is more plausible, though it loses force in view of Hunt's (1959) report of null results in a Cieutat-Stockwell-Noble (1958, Exp. I) type of experiment in which half the *S*s in each group articulated the paralogs while the other half did not. It is important to note that Hunt's material, the same dissyllables of the *m* scale used herein, can be pronounced faster than CVCs can be spelled. Of course, no one can be sure without experimenting that S-R familiarization effects do not interact with distributional factors, but the likelihood appears slim in the present case because *S*s in all groups except the Control were trained for 300 trials to pace their pronunciation responses at a 1-sec. rate during the familiarization phase. Furthermore, the S_0-R_0 group is slightly superior to both the Control and the S_0-R_{20} groups on R+ performance throughout the learning phase. The same phenomenon occurs in Cieutat's (1960) study. Both "spew" and distributional hypotheses imply the exact opposite of these observations. Appropriate experiments to settle this question will occur readily to most readers; meanwhile, the second author has some research in progress which indicates that type and tate of *n* are also relevant. The S-articulation factor is yet to be explored.

Our best judgment at the present writing is that no serious artifacts or procedural errors are contaminating the interpretation presented above. To repeat, however, the familiarization issue is not yet definitely settled. Much undoubtedly remains to be learned about the experimental control of *n* and the measurement of its effects.

SUMMARY

The role of familiarization (*n*) in paired-associate verbal learning was investigated. One hundred twenty *S*s, divided into five independent groups of 24 each, were used. Four experimental groups were formed by all combinations of locus (S vs. R term) and amount (0 vs. 20 exposures) of familiarization, while a Control group received no familiarization experience. All groups then practiced a five-unit paired-associate list of dissyllables of low initial familiarity (*f*) and meaningfulness (*m*) for 17 trials.

The influence of relevant S-term familiarization was positive and significant ($P < .01$), but not that of relevant R-term familiarization. Analyses of correct responses (R+) revealed no evidence of an irrelevant (learning-to-learn) familiarization effect, nor of any interactions between S or R and amount of practice. The initial effects of familiarization on incorrect responses (R−) were also significant ($P < .05$), causing decrements in the (irrelevant) S_0 groups and increments in the (relevant) S_{20} groups when compared with nonfamiliarized controls. After the first correct (reinforced) anticipation, acquisition rates of R+ and extinction rates of R− were uniform for all groups.

It was concluded that familiarization (*n*) and meaningfulness (*m*) operations do not produce equivalent effects on verbal behavior under these conditions. Contradictory results from other laboratories were discussed, and a tentative two-factor hypothesis was presented as an attempted synthesis of the differential roles played by *m* and *n* in paired-associate learning.

REFERENCES

ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269–276.

BAILEY, J. H., & JEFFREY, W. E. Response strength and association value in stimulus predifferentiation. *Psychol. Rep.*, 1958, **4**, 715–721.

CIEUTAT, V. J. Supplementary report: Stimulus and response meaningfulness (*m*′) in paired-associate learning by hospitalized mental patients. *J. exp. Psychol.*, 1959, **58**, 490.

CIEUTAT, V. J. Differential familiarity with stimulus and response in paired-associate learning. *Percept. mot. Skills*, 1960, **11**, 269–275.

Cieutat, V. J., Stockwell, F. E., & Noble, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m*, *m'*) in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

Hovland, C. I., & Kurtz, K. Experimental studies in rote-learning theory: X. Prelearning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.

Hunt, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, **57**, 262–267.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Mandler, G., & Campbell, E. H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 269–273.

Morikawa, Y. Functions of stimulus and response in paired-associate verbal learning. *Psychologia*, 1959, **2**, 41–56.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430. (a)

Noble, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446; **44**, 465. (b)

Noble, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.

Noble, C. E. The familiarity-frequency relationship. *J. exp. Psychol.*, 1954, **47**, 13–16.

Noble, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

Noble, C. E. Supplementary report: Familiarity and frequency. *J. exp. Psychol.*, 1960, **59**, 432–433; **60**, 418.

Noble, C. E. Verbal learning and individual differences. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961, in press. Ch. 6.

Noble, C. E., & McNeely, D. A. The role of meaningfulness (*m*) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

Parker, G. V. C., & Noble, C. E. Effects of experimentally-produced meaningfulness (*m*) on paired-associate learning. *Amer. Psychologist*, 1960, **15**, 451. (Abstract)

Riley, D. A., & Phillips, L. W. The effects of syllable familiarization on rote learning, association value, and reminiscence. *J. exp. Psychol.*, 1959, **57**, 372–379.

Robinson, E. S. *Association theory today.* New York: Century, 1932.

Sheffield, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctoral dissertation, Yale University, 1946.

Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Waters, R. H. The law of acquaintance. *J. exp. Psychol.*, 1939, **24**, 180–191.

Weiss, R. L. The role of association value and experimentally produced familiarity in paired associate verbal learning. (Unpublished doctoral dissertation, University of Buffalo, 1958) *Amer. Psychologist*, 1958, **13**, 386. (Abstract)

Winzen, K. Die Abhängigkeit der paarweisen Assoziation von der Stellung des besser haftenden Gliedes. *Z. Psychol.*, 1921, **86**, 236–252.

(Early publication received August 17, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 24–34

PAIRED-ASSOCIATES LEARNING WITH VARYING RELATIVE PERCENTAGES OF OCCURRENCE OF ALTERNATIVE RESPONSE MEMBERS[1]

ALBERT E. GOSS AND MARILYN E. SUGERMAN

University of Massachusetts

The purpose of the present experiment was to investigate learning with divergent paired-associates (PA) units: each divergent PA unit consisted of a single stimulus member and two response members, one of which occurred on $p\%$ and the other of which occurred on $(1\text{-}p)\%$ of the presentations of the stimulus member. One of the two major variables was relative percentages (RP) of occurrence of the alternative response members of each unit; the other was number of units.

Voss, Thompson, and Keegan (1959) have reported what are apparently the first experiments on acquisition with divergent units under conventional conditions of PA learning. In Exp. I and II, a different one of six RP from 50–50 to 100–0 in 10% steps was assigned in counterbalanced fashion to one of the units of a six-unit list. In Exp. III, the same PA list was used, but one of five RP from 50–50 to 90–10 was assigned to all six units. Whether the same *S*s had all or but one of the RP, percentages of anticipations of the more frequent response member for RP from 60–40 to 90–10 exceeded actual percentages of occurrence of those members. Anticipations of the alternative responses of the 50–50 RP were at or below 50%.

The present experiment complemented and extended those of Voss, Thompson, and Keegan. First, CVC nonsense trigrams of the present experiment were more conventional in form and origin than their CVCC units from a newly developed list. Second, the one-unit and four-unit lists of the present experiment filled the gap below their list with the same RP for all six units. Third, the one-unit and four-unit lists constituted a further variable: number of divergent units in a list, each with the same RP.

Number of units was introduced because of the general importance of length of material for learning and performance. The one-unit value was selected in part because it was the minimum length. More importantly, the stimulus-response relationships of a single divergent PA unit seemed homologous to those of certain simple two-choice situations for human *S*s involving a single first stimulus followed by one or the other of two second stimuli, each of which presumably evokes a different response.[2]

---

[1] Sponsored by the Office of Naval Research under Contract Nonr 2691(00). Susan Gregory helped to prepare materials for the experiment, Nancy Farrick assisted in the collection and analyses of data, and Alice Reilly aided in analyses of data.

[2] Simple two-choice situations which involve a single first stimulus followed only by one or the other of two second stimuli should be distinguished from those which involve a single first stimulus and not only one or the other of two second stimuli but also additional correlated stimuli (e.g., one amount of reward accompanying one of the second stimuli and another amount of reward or loss of reward accompanying the other of the second stimuli). Only findings and theory for the former situation are pertinent to learning with divergent PA units whose second stimuli are only the response members of a unit.

Accordingly, the one-unit list tested the generality of equations, such as Estes' (1959b, p. 411), whose implications for a single divergent PA unit with two response members are that percentages of anticipations of the more frequent (or the less frequent) response member will approach actual percentages of occurrence of that response member in negatively-accelerated fashion; at a rate independent of RP. Lists of *n* divergent PA units can be conceptually decomposed into *n* single units. The four-unit list, therefore, tested whether or not predictions based on equations and findings for simple two-choice situations and also for a single divergent unit could be extended without modification to lists of more than one unit.

METHOD

Experimental design.—Six relative percentages of occurrence of alternative response members of each PA unit (50–50, 60–40, 70–30, 80–20, 90–10, 100–0) were used with a four-unit list and with one-unit lists involving three conditions of presentations. (Hereafter the four-unit list and the three conditions of presentation are collectively referred to as lists, list conditions, or L.) The three conditions of presentation controlled for possible differences in learning with four-unit and one-unit lists attributable to length and variability of time intervals between successive occurrences of the same PA unit, and to the nature of the stimuli and *S*s' activities during those intervals.

With a four-unit list, each repetition of a particular unit is, on the average, separated by occurrences of the three other units. With a one-unit list, the time interval between each occurrence of the unit might be the same as the time interval between each of the four different units of a four-unit list. If so, the time interval between the single unit would be short and constant, while the time intervals between repetitions of the same unit in a four-unit list would be longer and variable. The *one-unit* condition involved a constant 2-sec. interval between successive presentations of the paired-associates unit. The *one-unit spaced* condition involved time intervals between successive presentations of the paired-associates unit whose average length and variability of the time intervals between successive presentations of that particular unit in a four-unit list. The one-unit spaced condition did not control for differences between four-unit and one-unit spaced lists with respect to rehearsal, associations with syllables of other units, and warm up. With the *one-unit random* condition, in order to minimize rehearsal of the responses of a particular unit and to provide warm-up experiences as well as the possibility of associations between the syllables of that particular unit and those of other units, time intervals between successive occurrences of the same paired-associates unit were filled by occurrences of pairs of syllables whose pairing was random rather than constant.

Subjects.—Men and women students enrolled in the introductory course in psychology at the University of Massachusetts participated as *S*s in partial fullfillment of course requirements. The 24 combinations of RP and L were run in eight counterbalanced cycles, with *S*s being assigned to a particular combination in order of its occurrence and *S*s' appearance until there were 8 *S*s in each combination, a total of 192 *S*s. Data from 28 other *S*s could not be used because of *E*s' errors ($N = 8$), apparatus failures ($N = 8$), or *S*s' inability of refusal to continue ($N=12$). Seven of the latter 12 were distributed among three RP with the four-unit list; the remainder were distributed among the three one-unit conditions.

Lists.—The CVC trigrams of the PA units of the four-unit and one-unit lists, with the recently obtained Archer (1960) association values in parentheses, were WOM (67%) followed by TEX (100%) or LIK (91%), RUF (86%) followed by DOL (94%) or CEN (75%), FES (69%) followed by VIC (97%) or BAL (91%), and NAV (72%) followed by HOB (90%) or SUR (91%). None has more than one letter in common with any other. The stimulus member and the one response member of each PA unit of a given trial were typed in elite capitals on the same line, 3 in. apart, on white paper, then attached to white plastic cards and presented by Hunter Card Master.

(*a*) *Four-unit.* Each trial consisted of one presentation of all four of the stimulus members, each of which was accompanied by one or the other of the two response members of that unit. Orders of occurrence of the four stimulus members during each trial, regardless of occurrences of response members, were determined by selecting two groups of 10 of the 24 permutations of four events. The

10 permutations of one group differed from those of the other group. Within each of these groups of 10 permutations, occurrence of a particular permutation was randomly determined. Since 120 trials were administered, each of the two groups of 10 permutations was repeated in a sequence which was determined randomly, subject to the constraints that each group of 10 permutations should occur 6 times and that neither group should occur more than 3 times in succession.

For the 50–50 RP, WOM was followed by LIK 50% of the time and by TEX 50% of the time, and similarly for the pairs of response members for RUF, FES, and NAV.

Under each of the other relative percentages, TEX, DOL, VIC, and HOB occurred with the same larger or smaller percentage and LIK, CEN, BAL, and SUR occurred with same smaller or larger percentage, respectively. For example, for half of the *S*s with RP of 60–40, TEX, DOL, VIC, and HOB each occurred on 60% of the trials while LIK, CEN, BAL, and SUR each occurred on 40% of the trials. For the other half of the *S*s TEX, DOL, VIC, and HOB each occurred on 40% and LIK, CEN, BAL, and SUR each occurred on 60% of the trials. Thus, relative percentages of occurrence of TEX, DOL, VIC, and HOB and of LIK, CEN, BAL, and SUR were counterbalanced. For the other RP, occurrences of members of these two sets of alternative responses were counterbalanced in the same way.

There were two constraints on the randomness of occurrences of alternative response members of each unit. One constraint was that the particular RP would be realized for each unit within each of the two groups of 10 permutations of orders of occurrence of stimulus members. The other constraint was that the patterns with which these RP were realized were the same through each of the six repetitions of each of the two groups of 10 permutations.

(*b*) *One-unit.* For each RP, there were eight one-unit lists, each of which was one of the units of the four-unit list with one or the other response member occurring more frequently. For the 50–50 RP, the order of the alternative response members was interchanged. Each list was administered to a different *S*.

(*c*) *One-unit spaced.* Length and variability of time intervals between successive presentations of the unit were the only differences between one-unit and one-unit spaced conditions. These intervals were obtained for each of the eight lists by inserting blank cards for the three PA units of the four-unit list which were not being used.

(*d*) *One-unit random.* Instead of the blank cards of the one-unit spaced condition there were pairs of syllables which had been formed by drawing 30 successive pairs of syllables at random from decks of 60 syllables. Each deck consisted of the nine syllables which were the members of the three units not being used as the constant unit. There were 10 instances of each of three syllables which were stimulus members of these three units and five instances of each of the six syllables which were response members. Within the 30 pairs, each of the nine syllables was paired with each of the other eight syllables, and each was, on the average, a stimulus member half of the time and a response member half of the time.

Apparatus and procedures.—Each unit was presented at a 2:2 sec. rate by means of a Hunter Card Master. Each exposure of the blank cards of the one-unit spaced lists was for 4 sec. The interunit interval of the four-unit, and one-unit random conditions was 2 sec. The interunit interval and the intervals between units and blank cards or between blank cards of the one-unit spaced condition were each 2 sec., but counting the time required for occurrences of blank cards, time intervals between successive presentations of the single unit ranged from 2 to 38 sec. about a mean of 20 sec.

Each of the units of the four-unit list and the single unit of the one-unit, one-unit spaced and one-unit random conditions was presented 120 times. Under the one-unit spaced condition, the 120 presentations of a unit alternated with 360 presentations of blank cards. Under the one-unit random condition, the 120 presentations of a unit alternated with 360 presentations of the randomly-formed pairs.

The same essentially standard instructions for paired-associates learning, with anticipations spelled, were given to *S*s in each of the 24 combinations of conditions. They were not told that each stimulus member might have alternative response members.

RESULTS

For RP from 60–40 to 100–0, the response measure was percentage of anticipations of the more frequent and less frequent response member of each unit in each of six successive blocks of 20 trials each. During the last 60 trials under all four list conditions and essentially for all trials with the

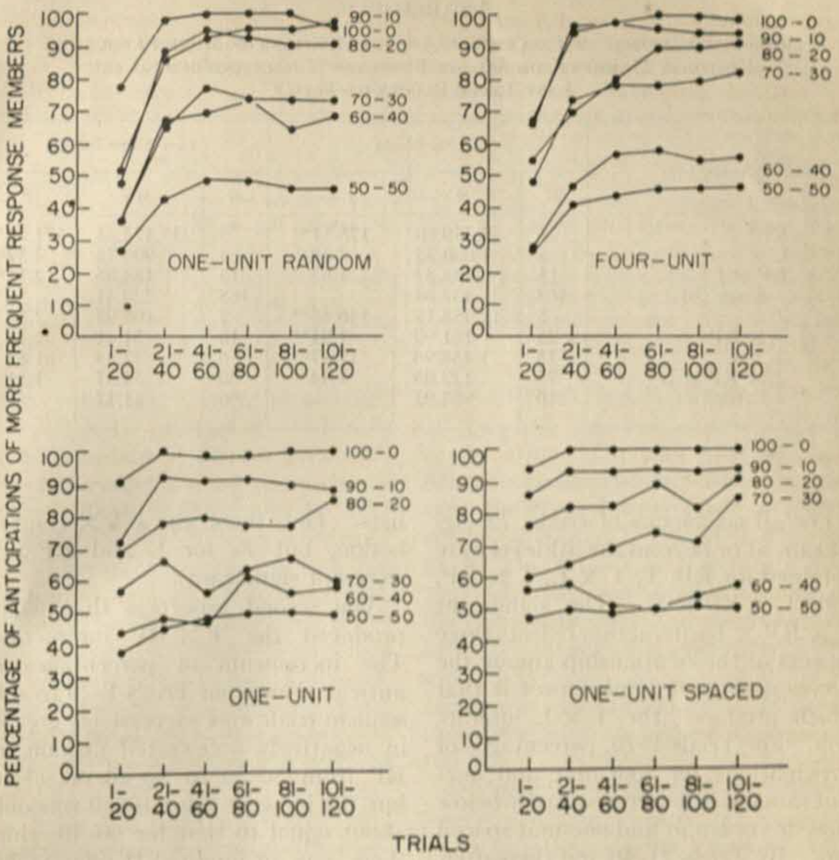


FIG. 1. Mean percentages of anticipations of the more frequent response members for successive 20-trial blocks for each of the 24 combinations of RP and L.

one-unit spaced and one-unit random lists percentages of anticipations of the more frequent and less frequent response members for each percentages summed almost to or to 100%. For this reason only percentages of anticipations of the more frequent members are presented. For the 50-50 RP, the measure for each unit was the average of the percentages of anticipations of both response members of the unit in each successive block of 20 trials. For each *S* administered the four-unit list, percentages of anticipations of response members for each of the four units were averaged to obtain a single percentage for each block.

Over-all differences among combinations of conditions.—Figure 1 shows means of percentages of anticipations of more frequent response members for successive blocks for each of the 24 combinations of RP and Lists (L). Differences among these trends for all six and during the last three blocks of trials (T) were assessed by the analyses of variance summarized in Table 1.[3]

---

[3] With the 100-0 relative percentages, *Ss* responded almost perfectly from Trials 21-40 on. In order to save both some *Ss*' and *Es*'

TABLE 1

Analyses of Variance of Percentages of Anticipations of More Frequent Response Members for All Six Blocks of Trials and during the Last Three Blocks of Trials

Source	All Six Blocks			Last Three Blocks		
	df	*MS*	*F*	*df*	*MS*	*F*
RP	5	77,230.60	178.51**	5	35,422.54	171.84**
L	3	1,030.33	2.38	3	990.24	4.32**
RP × L	15	723.33	1.67	15	483.86	2.11*
Error (b)	168	432.63		168	229.41	
T	5	12,458.15	146.85**	2	105.03	2.53
T × RP	25	161.90	1.91**	10	51.41	1.24
T × L	15	1,458.98	17.17**	6	77.38	1.87
T × RP × L	75	122.08	1.44**	30	54.37	1.31
Error (w)	810	84.95		306	41.47	

* $P < .05$.
** $P < .01$.

For all six blocks of trials, *F*s significant at or beyond the .01 level were obtained for RP, T, T × L, T × RP, and T × RP × L. The significant T × RP × L interaction reflects three aspects of the relationship among the curves of Fig. 1. One aspect is that which produced the T × L interaction. For Trials 1–20, percentages of anticipations for four-unit and one-unit random lists were markedly below those for one-unit and one-unit spaced lists. By Trials 21–40 and thereafter, the percentages for the former two lists were slightly above those for the latter two lists. When averaged across all 120 trials, the early lower and subsequent higher values of the percentages for four-unit and one-unit random lists yielded percentages across all 120 trials which were about the same as those across all 120 trials for the one-unit and one-unit spaced lists. Thus there was a T × L interaction, but *F*s for L and RP × L were not significant.

The second aspect is that which produced the T × RP interaction. The increments in percentages of anticipations from Trials 1–20 to subsequent trials were successively greater in negatively accelerated fashion for RP from 50–50 to 80–20 or 90–10, but the increment for 100–0 was only about equal to that for 60–40; thus, there was an inverted-U relationship between these increments and RP. The third aspect is that, unlike the curves for RP from 60–40 to 90–10, the curves for 50–50 for four-unit and one-unit random lists tended to be slightly below those for one-unit and one-unit spaced lists even during the last 60 trials. For 100–0, all four curves were about equal from Trials 41–60 onward.

The *F* for RP reflects a direct relationship between percentages of anticipations and actual percentages of occurrence which held not only across but also within combinations of L and T.

For the last three blocks of trials, neither the *F* for T nor *F*s for inter-

---

time, the last 40 trials were eliminated for 15 *S*s who were under 100–0 RP with four-units, one-unit spaced, and one-unit lists. For these three combinations, the *N*s for those blocks of trials were reduced by 4, 6, and 5, respectively. Missing values for the last two blocks were estimated from cell means; the *df*'s of "within *S*s" error variances were reduced by 30.

TABLE 2

MEANS AND *SD*s OF PERCENTAGES OF ANTICIPATIONS OF MORE FREQUENT RESPONSE MEMBERS DURING TRIALS 61–120

Lists (L)	Relative Percentages of Occurrence (RP)											
	50–50		60–40		70–30		80–20		90–10		100–0	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Four-unit	46.5	2.1	56.4	11.6	80.0	10.0	91.0	5.6	95.3	2.5	99.5	0.4
One-unit random	47.0	4.4	69.6	19.5	74.5	15.5	91.6	5.5	96.9	3.6	99.2	2.3
One-unit spaced	49.8	0.4	53.8	9.4	76.5	7.8	87.3	7.9	94.0	4.4	100.0	0.0
One-unit	49.8	0.3	58.7	6.6	63.0	13.9	80.0	7.6	90.2	9.7	99.8	0.2

actions involving T were significant. The RP × L interaction indicates that rank-orders of the four list conditions, differences among those conditions, or both differed among RP. For 50–50 and 100–0, there were only slight differences among the four list conditions. For the remaining RP, with some inversions, the rank-order of decreasing percentages of anticipations was one-unit random, four-unit, one-unit spaced, and one-unit lists. The differences among list conditions for these RP were sufficient to produce the significant *F* for L. The significant *F* for RP reflects a direct relationship between percentages of anticipations and actual percentages both across and within list conditions.

Since there were no further changes in mean percentages of anticipations of the more frequent response members during the last 60 trials, means of percentages of anticipations for Trials 61–120 were computed as single, more reliable estimates of asymptotic percentages (Table 2). Rates of approach (θ') to these presumed asymptotes were estimated by the equation derived by Voss (1958).[4] Table 3 presents the values of θ' obtained for all but the one-unit spaced, 60–40 combination. (Because of a decrease between earlier and terminal percentages, the θ' for this combination could not be computed.) For each RP, the θ's for the three one-unit conditions were, with two exceptions, larger than the corresponding θ' for the four-unit list. Within each list condition, with some inversions, values of θ' first decreased from those at 50–50 and then increased to highest values at 100–0.

Differences among specific combinations of conditions.—Mean percentages of anticipations during Trials 61–120 were also used for comparisons among specific combinations of conditions.

TABLE 3

VALUES OF θ' FOR COMBINATIONS OF RP AND L

Lists (L)	RP					
	50–50	60–40	70–30	80–20	90–10	100–0
Four-unit	.085	.075	.102	.067	.194	.136
One-unit random	.113	.100	.087	.109	.085	.234
One-unit spaced	.678	—	.106	.204	.470	.800
One-unit	.131	.079	.296	.305	.305	.544

---

[4] The equation used was

$$\theta' = p(\infty)/np(\infty) - \sum_{1}^{n} \overline{p(n)}$$

where θ' is the estimate of learning rate, $p(\infty)$ is the estimated asymptote of percentages of anticipations of the more frequent response member, n is the number of trials, and $\sum_{1}^{n} \overline{p(n)}$ is the total number of occurrences of the more frequent response in n trials.

TABLE 4

Differences (*d*) between Percentages of Anticipations of More Frequent Response Members and Their Actual Percentages of Occurrence during Trials 61–120 and *t*'s for Those Differences

Lists (L)	Relative Percentages of Occurrence (RP)									
	50–50		60–40		70–30		80–20		90–10	
	d	*t*	*d*	*t*	*d*	*t*	*d*	*t*	*d*	*t*
Four-unit	−3.5	4.38**	−3.6	0.82	+10.0	2.64*	+11.0	5.16**	+5.3	5.64**
One-unit random	−3.0	1.86	+9.6	1.30	+4.5	0.77	+11.6	4.11**	+6.9	5.06**
One-unit spaced	−0.2	0.48	−6.2	1.76	+6.5	2.22	+7.3	2.44*	+4.0	2.45*
One-unit	−0.2	1.39	−1.3	0.53	−7.0	1.33	0.0	0.0	+0.2	0.07

* $P < .05$, $df = 7$.
* $P < .01$, $df = 7$.

Mean percentages for 100–0 were excluded from these comparisons to reduce heterogeneity of variances, and to eliminate the problem of some estimated percentages over the last 40 trials. Also, the 100–0 RP differed from the others: while the 100–0 could only be below or at 100%, the others could be below, at, or above the percentages of occurrence of the more frequent response member.

A Duncan multiple range test (Federer, 1955, pp. 26–29) indicated that for RP of 50–50, 80–20, and 90–10, there were no differences among the four lists conditions. For 60–40, the percentage for the one-unit random list was significantly larger than that for the one-unit spaced list. For 70–30, the percentage for the one-unit list was significantly smaller than the percentages for one-unit spaced and four-unit lists. In general, percentages of anticipations for a particular list condition within each RP did not differ significantly from percentages of anticipations for that condition within RP 10% removed.

Percentages of anticipations compared with percentages of response-member occurrence.—For each RP except 100–0, Table 4 presents *t*'s for differences between mean percentages of anticipations of more frequent response members and actual percentages of occurrence of those members. The error terms of these *t*'s were computed from the *SD* for each particular combination. Since Wilkinson's (1951) table indicates that eight of 20 *t*'s significant at the .05 level would occur less than once in a 1000 times, the significant *t*'s as a whole cannot readily be attributed to chance factors. For the four-unit list with RP of 50–50, the percentage of anticipations was significantly lower than 50%. All other significant *t*'s are for percentages of anticipations of more frequent response members which exceeded their actual percentages of occurrence. For the four-unit list, differences for 70–30, 80–20, and 90–10 were significant at from .05 to less than .01. For both one-unit spaced and one-unit random lists, the differences for 80–20 and 90–10 were significant at from .05 to less than .01. No difference for the one-unit list approached significance.

Differences between sets of more frequent response members and among units of the four-unit list.—Within the four-unit list, TEX, DOL, VIC, and HOB occurred both more frequently and less frequently than their respective alternative response members, LIK, CEN, BAL, and SUR. An analysis of variance revealed no difference between the two sets of more frequent alternative response members.

During the first three blocks of trials, analyses of variance indicated that there were differences among units for both more frequent and less frequent occurrence of TEX, DOL, VIC, and HOB. However, during the last three blocks, the units were equivalent.

DISCUSSION

Three features of the results of this experiment are of particular importance. One is the effects of RP and L on rates of approach to and apparent asymptotes of percentages of anticipations of more frequent response members. The second is implications of the present results for the generality of theory and findings for two-choice situations with light stimuli and a single unit. The third is the possible use of divergent paired-associates units to establish hierarchies of stimulus-response associations.

Percentages of anticipations of the more frequent response member.—Over the last three blocks of trials, none of the *F*s involving trials was significant. Thus, for each of the 24 combinations of RP and L, asymptotic percentages of anticipations were apparently reached on or before Trials 61–80. Only with certain combinations of RP and L did asymptotic percentages of anticipations of the more frequent response equal their actual percentages of occurrence.

For 50–50, with both one-unit and one-unit spaced lists, the asymptotic percentages of anticipations were almost exactly 50%. But the estimated asymptote of 46.5% with the four-unit list was significantly below 50%, and the decrease from 50% with the one-unit random list, though not statistically significant, was almost as great. The remaining responses were sometimes intralist intrusions which indicates that these apparent decrements were largely due to interference from responses to the stimuli of other units of the four-unit and one-unit random lists.

For 60–40, the smallest difference between percentages of anticipations and actual percentages was with the one-unit list. Even though the absolute values of the other three differences were up to 9.6%, none was statistically significant. With the one-unit list, only with RP of 70–30, was there a sizeable difference between percentages of anticipations and actual percentages. But this difference, in the direction of fewer than 70% anticipations of the more frequent response member, was not statistically significant. The other three differences represented anticipations of the more frequent response member exceeding 70%, but only that for the four-unit list was statistically significant.

For 80–20 and 90–10, with the one-unit list the percentages of anticipations of the more frequent response were almost exactly equal to 80% and 90%. But all of the percentages of anticipations with the other three lists were significantly larger than 80% and 90%. Thus, for all but the one-unit list, there is apparently some point between 60–40 and 80–20 at which percentages of anticipations of the more frequent response member begin to exceed their actual percentages of occurrence. Occurrences of the alternative responses of each unit were semirandom rather than random, a circumstance which should facilitate matching actual percentages. Accordingly, those higher percentages of anticipations than of actual percentages which were obtained may represent minimum rather than maximum differences.

For almost all RP, percentages of anticipations with the four-unit and one-unit random lists seemed more similar to each other than to those with the one-unit spaced list which differed somewhat from the one-unit list. The former difference suggests that, whether the pairings of the stimuli between a particular unit were constant or random, the stimulation and activities associated with occurrences of those units had some differential effect on anticipations relative to repetitions of a particular single unit separated by the same longer, variable intervals filled with occurrences of blank cards. For RP of 80–20 and above, the longer, variable intervals seem to occasion percentages of anticipations greater than the matching percentages obtained with the one-unit list.

For 70–30, 80–20, and 90–10 with four-unit, one-unit random, and one-unit spaced lists, the results of the present experiment agree closely with Voss et al. (1959) findings of percentages of anticipations of more frequent response members larger than their actual percentages of occurrence. For these three list conditions, which are more similar to Voss et al. lists than is the one-unit list, there is only one pronounced deviation from their results: for 60–40, percentages of anticipations with four-unit and one-unit spaced lists were below rather than above 60%. This disagreement may reflect chance factors, and hence is only apparent, or it may represent replicatable effects of one or more differences in stimuli, apparatus, and procedures.

Generality of equations.—Further use of equations such as Estes' (1959b, p. 411) for predictions of acquisition with divergent PA units, without some modification, apparently depends on the particular combinations of RP, number of units, and conditions of presentation of one-unit lists. With the one-unit list, asymptotic percentages for all RP and actual percentages did not differ significantly and only with 70–30 was the value of the difference of any magnitude. Thus, the equation seems applicable without modification. However, this relatively close fit of anticipations and actual occurrences has one perplexing feature: a closer fit would probably have been predicted with the longer, variable interunit intervals of the one-unit spaced list than with the short, interunit interval of the one-unit list (1959b, p. 421), while the results for the two lists were just the opposite. Conceivably the two conditions would not have had different effects had *S*s merely waited during the longer, variable intervals rather than looking at blank cards.

With the one-unit spaced and one-unit random lists for RP of 80–20, 90–10, and perhaps of 70–30, and with the four-unit list for RP of 70–30, 80–20, and 90–10, asymptotic percentages of anticipations of the more frequent member exceeded actual occurrences. Furthermore, with the four-unit list and perhaps the one-unit random list for RP of 50–50, asymptotic percentages were apparently less than 50%. Thus, any satisfactory equation must provide for predictions of asymptotic percentages of more frequent response members which, for specific combinations of RP and L, are less than, equal to, or greater than actual percentages of occurrence of those members. For this objective, even more recent extensions based on "component and pattern models" (Estes, 1959a) seem inadequate. The rate parameter also seemingly varied with RP and L in a manner roughly consistent with Voss et al. findings. Therefore, either values for this parameter must be determined empirically for each specific combination of RP and L, or an equation expressing θ' as a function of RP and L must be developed.

As predicted by equations and findings for two-choice tasks involving simpler stimuli and a single unit, the curves for most of the combinations of RP and L were negatively accelerated. The exceptions were some for RP of 50–50, 60–40, and 70–30 with one-unit and one-unit spaced lists whose initial negative acceleration was completed so early in the first 20 trials, that their means for those trials were essentially at asymptotic percentages.

Hierarchies of stimulus-response associations.—The divergent PA unit may prove useful for experimentally establishing hierarchies of associations of two or more responses to what is ordinarily conceived as the same or a single stimulus. (Analytically, such stimuli may best be treated as constituted of n elements or as appearing in one or more of n patterns each of which is associated with one response in all-or-none fashion.) At present, hierarchies of stimulus-response associations or of responses are usually inferred from hierarchies of responses to a stimulus for large groups of *S*s. On the assumption of the existence of hierarchies of responses to a single stimulus, a procedure for establishing such hierarchies on the part of individual *S*s would have two advantages: (*a*) inferences from hierarchies for groups to those for individual *S*s could be replaced by more direct and presumably more precise specification of hierarchies for individual *S*s, and (*b*) with different hierarchies of responses to the same stimulus, confounding of different stimuli with different hierarchies could be eliminated.[5]

Following a suggestion by Hull (1930, pp. 244–245), Peterson (1956) tested a less direct procedure for establishing hierarchies of responses to stimuli. The stimulus members of Peterson's paired-associates were compounds of two nonsense syllables. One response was conditioned to one syllable of the compounds and two responses were conditioned to the other syllable of the compounds. After the learning criterion has been attained, the syllables of the compounds were presented separately to permit determination of frequencies of alternative responses to the syllable to which those responses had been conditioned. A disadvantage of this procedure is that acquisition of potential hierarchies cannot be observed except by interpolated test trials. Also, although Peterson was able to predict observed responses from his formula, the procedure involves the possible complication of adjustments for training with a stimulus in a compound and testing with the stimulus alone (Hull, 1943, p. 220).

Osgood and Anderson's (1957) task could be viewed as consisting of 10 paired-associates units each of which had three response members which occurred with percentages of .60, .30, and .10. However, since each name was both a stimulus member and a response member, because of "backward associations," it is difficult to specify the potential number of responses in a hierarchy. Thus, the name "Josephine" of Osgood and Anderson's illustration might be regarded as having not three but six different responses.

SUMMARY

Acquisition with divergent paired-associate (PA) units was investigated as a function of six relative percentages of occurrences of alternative response members of a unit (RP) which varied from 50–50 to 100–0 in 10% steps, and of number of units, four or one. In order to control for effects of time intervals between presentations of the same unit in the four-unit list and for events within those intervals, there were three conditions of presentation of single units: (*a*) a constant, 2-sec. interval between each presentation; (*b*) variable, longer intervals between each presentation, and (*c*) the same variable, longer intervals filled with presentations of randomly-paired stimuli. Eight undergraduate *S*s were assigned to each of the 24 combinations of RP and L wherein they had 120 trials with each unit of the four-unit list or with the single unit of the three one-unit conditions.

Across all six blocks of 20 trials each, and during the last three blocks, there was a direct relationship between percentages of anticipations of the more frequent response members and their actual percentages of occurrence. During the last three blocks, decreasing percentages of anticipations were obtained with one-unit random, four-unit, one-unit spaced, and one-unit lists. Differences in rate of acquisition were reflected in significant interactions of the six blocks of trials with RP, with L, and with RP and L.

The RP × L interaction for the last three blocks of trials could be interpreted as reflecting differences among lists with respect to the degree to which percentages of anticipations of the more frequent response members approximated their actual percentages of occurrence. With the one-unit list, percentages of anticipations were essentially the same as actual percentages. With four-unit, one-unit random, and one-unit spaced lists for 80–20 and 90–10, and probably for 70–30, percentages of anticipations exceeded actual percentages.

---

[5] Estes (1960, p. 212) has recently presented analyses of responses after two or four trials with divergent PA units the pattern of which was interpreted as evidence against an assumption of an hierarchy of responses to a single stimulus for individual *S*s. Until data for test-trial performance following more than one or two occurrences of the alternative response members are available, both the generality of those findings and his interpretation seem questionable. Examination of the data for the first 20 trials for individual *S*s of the four-unit 50–50 combination of the present experiment disclosed that, in 18 of 32 instances, one or the other response member was anticipated on 75% of the occurrences of a particular unit. Thus, Estes may simply have demonstrated that one or the other of the alternative stimulus-response relationships of a divergent unit, presumably because of pre-experimental experiences, has considerably and consistently greater initial strength.

The significance of these results for findings and theory for apparently homologous situations involving simpler stimuli was discussed. Also noted was the possible use of divergent PA units for establishing response hierarchies experimentally.

References

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. (a)

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. (b)

Estes, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

Federer, W. T. *Experimental design: Theory and application.* New York: Macmillan, 1955.

Hull, C. L. Simple trial and error learning: A study in psychological theory. *Psychol. Rev.*, 1930, **37**, 241–256.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Osgood, C. E., & Anderson, L. Certain relations among experienced contingencies, associative structure, and contingencies in encoded messages. *Amer. J. Psychol.*, 1957, **20**, 411–420.

Peterson, L. R. Prediction of response in verbal habit hierarchies. *J. exp. Psychol.*, 1956, **51**, 249–252.

Voss, J. F. Supplementary report to "language behavior and subjective probability." *Nat. Inst. Ment. Hlth. Rep.*, 1958, No. M-1476.

Voss, J. F., Thompson, C. P., & Keegan, J. H. Acquisition of probabilistic paired associates as a function of $S\text{-}R_1$, $S\text{-}R_2$ probability. *J. exp. Psychol.*, 1959, **58**, 390–399.

Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156–158.

(Received April 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 35–42

SEPARATING PERCEPTUAL AND LINGUISTIC EFFECTS OF CONTEXT SHIFTS UPON ABSOLUTE JUDGMENTS[1]

DAVID L. KRANTZ AND DONALD T. CAMPBELL

Northwestern University

In accordance with the usual strategy of parsimony in science, the effects of extreme contexts upon absolute judgment were interpreted initially as illustrating a *single* process of relativity of judgment or adaptation level. The most economical interpretation of the data to be presented here necessitates the postulation of *two* such processes, both producing contrast effects in the same direction. One of these processes, transient and reversible, is found in both of two judgmental languages, and is designated *perceptual*. The other, irreversible and specific to an unfamiliar, experimenter-defined response language, is designated *semantic*.

Two papers published in 1958 independently suggest dividing the judgmental phenomena currently grouped under Helson's concept of adaptation level (e.g., Helson, 1947, 1948, 1959) into semantic or linguistic effects on the one hand, and perceptual or end-organ effects on the other. One of these papers (Campbell, Lewis, & Hunt, 1958) argued that the "situationally relative," "novel," "arbitrary" and "restricted" response language ordinarily used in the method of single stimuli could well have produced the contrast effects resulting from extreme anchors or extreme contexts without generating distortions in identity judgments (had they been called for) or any perceptual illusion. That study then examined the effect of a shift in context of stimulation upon the judgment of a common stimulus, employing a judgmental language which was "absolute, extensive, and extra-experimentally anchored." Finding the usual contrast or adaptation level effects, the paper concluded that such phenomena could not be explained away as semantic artifacts. However, it did not raise the possibility that both semantic and perceptual effects are present in the typical judgmental assignment.

The other paper suggesting a duality of processes was that of Stevens (1958) in which he argued that different phenomena were involved in sensory adaptation (in which end-organ receptor-cell adaptation takes place) on the one hand, and in judgment (where redefinition of categories is involved) on the other, although previous approaches treat the two as if the same process were involved. Here again, two processes were postulated with no explicit assumption that both could be present in the same judgmental task. The outcomes of the present study seem to require such an assumption.

The setting of the present study is the judgment of the lengths of white lines projected on a screen. A common "tracer" stimulus recurs in all phases of the experiment, in one phase interspersed among longer lines, in another phase occurring among shorter lines. One set of *S*s judges in "inches," the other uses an *E*-defined response language for which the value "100" is "average." The effects of the shift in context upon judgments of the common tracer stimulus are examined.

---

[1] This study is based upon an MA thesis done by the first author under the second author's direction. Certain analyses beyond those reported in the thesis (Krantz, 1959) have been financed by ONR Project 7onr-450(11), W. A. Hunt, Principal Investigator. The authors are also indebted to Allen Parducci for editorial suggestions.

METHOD

Slides were prepared which projected white lines of 13 different horizontal lengths, for a .2-sec. duration, with a 2-sec. interstimulus interval. The slides were made by using black tape on clear glass $3\frac{1}{4} \times 4$ in. slides. Projected on the screen, the shortest line (No. 1) was approximately 6 in., the longest 36 in., the middle "common" or "tracer" stimulus (No. 7) 20 in. (the exact values of the stimuli are presented in Table 1) and the vertical widths $3\frac{3}{4}$ in. The slides were organized in groups of seven (here called "trials") within each of which the common tracer Stimulus 7 occurred once, the order of stimuli within a trial being random. For the Low Phase, the stimuli in each trial ranged from Stimulus 1 to Stimulus 7. For the High Phase, the range was from Stimulus 7 to Stimulus 13. Trials in the Transition Phase also spanned seven steps, graduating between the Low and High Phases. The three transition trials ranged between Stimuli 3–9, 4–10, 5–11 for Low to High and vice versa for a High to Low transition. Both a High-Low and a Low-High group were run for each combination of the other three experimental variables.

In the second experimental treatment variable, the "Inches" condition was presented to *S*s with these instructions, "You are to judge these lines in terms of how long, in inches, they appear to be." The "Rating" condition was introduced this way:

> You are to judge these lines in terms of a rating scale which you will construct during the course of the experiment. When you see a line which appears "average in length" to you, judge that line with the number 100. If a line appears less than average, assign the number 95. If another line appears even smaller than the average line and smaller than the 95 one, use the number 90 to designate "less, less than average."
>
> In the same way, if a line is above average in length, assign the number 105. If another line is larger than the 100 average line and larger than the 105 line, assign the number 110. In other words, you can build your rating scale above and below the average rating of 100. You are not restricted to the numbers I have given you. The numbers I have given you on the reference sheet are only examples. In-between numbers, like 93 or 106 are appropriate. You can use numbers less than 90 or greater than 110. In other words, you are constructing a scale which shows the differing lengths of lines presented. Any system of numbers, such as in-between numbers or numbers above 110 or below 90 are usable. The only stipulations are that you use 100 as average and that if you assign a number to a particular length of line, the next time that line is presented you assign it the same number. Let me repeat, always use the same number for the same length of line; do not change your judgment of the same line.
>
> Let me review the categories or numbers on your reference sheet. The number 100 stands for the average line; 95 for less than average; 90 for less, less than average; 105 for more than average; 110 for more, more than average.

In choosing this particular form of rating, there was an effort to eliminate the effects of "vocabulary exhaustion" (Campbell, Lewis, & Hunt, 1958) such as occurs when all of a restricted number of ratings have been employed prior to a shift of context, and to exclude a deliberately relative or comparative usage by *S*s which would allow them consciously to redefine values in the course of the experiment.

The third experimental variable employed was Preshift Training, i.e., the length of the initial phase, presumably manipulating the persistence of the adaptation level established therein. "Phase 1" indicates the initial set of trials given, irrespective of the context employed. This period was followed by the transition series ("Phase 2"). A final series of trials after shift of context is designated "Phase 3." In the Short condition, five Phase 1 trials were used; in the Long condition, 15.

The fourth variable was the degree of illumination, which was presumed to affect stimulus discriminability. Under the Bright condition, the projected lines registered .73 apparent ft-c. Under the Dim condition a neutral density filter reduced the brightness to .32 apparent ft-c. The background brightness of the screen was .26 apparent ft-c for both illumination conditions.

These four dichotomies generate 16 experimental treatments. For each of these, 20 *S*s were obtained, a total of 320 *S*s. The *S*s were drawn from introductory classes in psychology. They were tested in groups of four to six, seated in two rows, 13 and 16 ft. from the screen. Experimental sessions were randomly assigned to treatments, with the restriction that all 16 treatments be evenly represented during all periods of the three month course. In two successive quarters the experimental design was replicated with 10 *S*s in each cell.[2]

The *S*s recorded their own judgments in a

---

[2] One session of 6 *S*s (in treatment High-Low, Rating, Short training, Dim illumination) was discarded because of extremely aberrant results. This session was subsequently rerun twice with different *S*s to confirm the aberrance, attributable to intraexperimental history (Campbell, 1957). It is felt that this substitution of the rerun group data augments rather than reduces the replicability of the study's findings.

TABLE 1

MEAN JUDGMENT OF EACH STIMULUS DURING FIRST FIVE TRIALS OF PHASE 1

Stimulus Number	Low Context			Stimulus Number	High Context		
	Length (In.)	Mean Judgment: Inches	Mean Judgment: Ratings		Length (In.)	Mean Judgment: Inches	Mean Judgment: Ratings
1	6.1	3.74	79.90	7	20.2	12.74	90.98
2	9.0	5.37	88.99	8	23.0	14.38	94.00
3	11.0	6.75	92.48	9	25.0	16.95	97.78
4	12.4	7.80	96.09	10	28.1	20.90	101.74
5	15.8	10.77	102.55	11	31.3	23.32	103.72
6	18.7	13.32	106.29	12	33.1	27.16	108.21
7	20.2	17.21	111.77	13	35.8	31.59	112.63
Mean		9.28	96.87	Mean		21.01	101.29

Note.—Preshift training and discriminability conditions pooled, $N = 80$ *S*s per mean.

booklet of slips with each page allowing 10 judgments. Twelve *S*s were discarded for getting out of step with the stimulus presentation and for incomplete records. Excess *S*s over the 20 required for each treatment were discarded by the use of random numbers.

RESULTS AND DISCUSSION

While the statistical analyses to be presented have been done separately for each judgment language, in Fig. 1 and 2, there is an effort to show both sets of results on the same graph. This necessitates computing a conversion factor between the two judgment languages, which in turn necessitates examining data on judgments of stimuli other than Stimulus 7, data not otherwise used. Because of their relevance to the attempted translation between the two languages, and to the more general problem of the relationship of stimulus and judgmental distances, these data are presented in Table 1.

The data do not lead to any univocal conversion ratio. If ratings be plotted against stimuli, the plots are satisfactorily linear, but of differing slope for the two contexts, the High context being the flatter. If judged inches be plotted against stimuli, the slopes depart slightly from linearity in the unexpected direction of positive acceleration for both High and Low contexts, with the High context being the steeper in this case. When rating means are plotted against judged inches, the Low context shows a slight bowing, and a steeper slope than the High context. As an approximate average of the slope of the two contexts, a ratio of 1 in. = 1.50 rating points has been chosen for the purpose of graphing Fig. 1 and 2. The graphs have been centered on the mean for the two contexts of the column means of Fig. 1, i.e. 15.64 in. and 99.08 rating points. Because of the differing slopes for the two contexts, this represents a more arbitrary decision than had been hoped for.

In Fig. 1 and 2, all data represent judgments of the same tracer stimulus, No. 7, and all connected points represent judgments by the same 40 *S*s. From the Phase 1 differences between the corresponding High-Low and Low-High groups, one can see that the expected contrast effects emerge. The striking shifts in each line between Phase 1 and Phase 3 are likewise uniformly in the expected direction. But perhaps most dramatic is the difference in overall pattern between the Inches and Ratings conditions, a difference which provides the central problem of interpretation.

A summary of the results of six independent analyses of variance, shown in Table 2, provides statistical

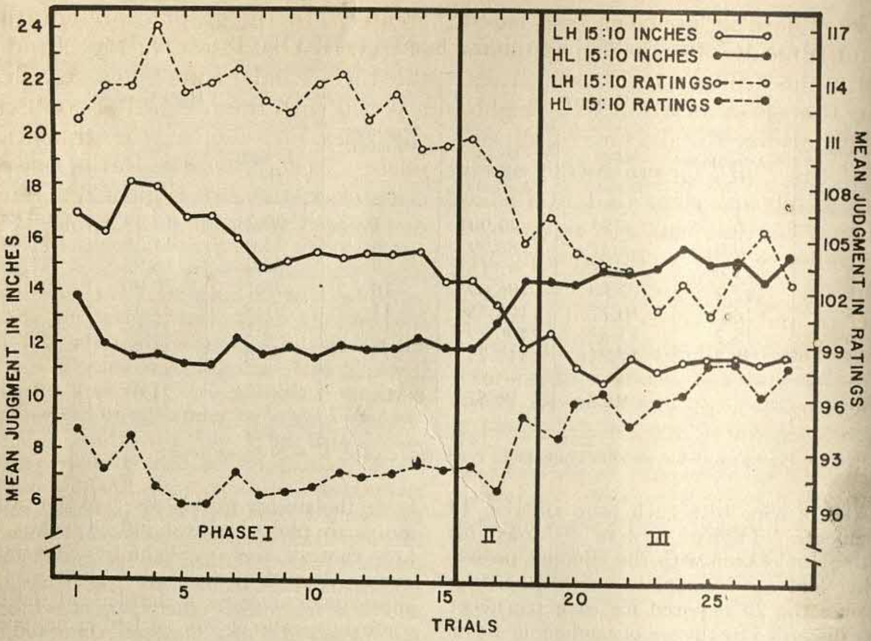


FIG. 1. Mean trial-by-trial judgment of tracer Stimulus 7 for Long training condition. (Discriminability conditions pooled, $N=40$ for each point.)

confirmation of the overall pattern for Inches and Ratings conditions. In the Phase 1 analyses, the mean of each *S*'s judgment of Stimulus 7 during the last four trials of Phase 1 has been used, since this set of responses is comparable across Preshift Training conditions. For both response languages, the main effect of context is highly significant and is the only significant effect. As graphically illustrated in Fig. 1 and 2, the effect is greater for the Ratings. Discussion of outcomes such as these frequently imply that the difference between the F ratios is itself significant. Rarely, however, is this higher-order significance tested. Bradley and Schumann (1957) have recently provided an F test among Fs appropriate for parallel experiments such as the present one. By their criteria ($w' = 21.32$, $a' = 33.47$, $b = 76$, $P < .05$, minimum $w'_{.05} = 1.85$),[3] the effect is significantly greater for the Ratings condition. This is one of the major findings of this study.

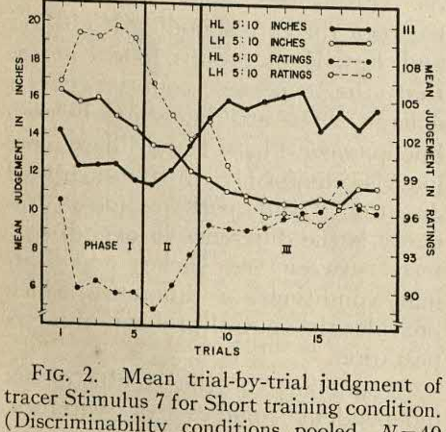


FIG. 2. Mean trial-by-trial judgment of tracer Stimulus 7 for Short training condition. (Discriminability conditions pooled, $N=40$ for each point.)

[3] The tables supplied by Bradley and Schumann provide w' values only at 5% level of significance. These minimum w' values will be presented in parenthesis following the obtained values. Note that the obtained values greatly exceed these minimums.

The Phase 3 analyses are based upon the means of each *S*'s judgments of Stimulus 7 in the 10 Phase 3 trials. Here the effect of context is highly significant for the Inches condition. (That the *F* of Phase 3 Inches seems more significant than that of Inches in Phase 1 is due primarily to reduced variability which is presumably attributable to the greater amount of practice and to scores based upon 10 judgments rather than 4.) As seen in Fig. 1 and 2, the effect of context on Ratings is weak and in the opposite direction from that on Inches. This main effect ($P < .05$) is graphically clear only for the Long Preshift Training. The difference in Phase 3 context effects *F*s between Ratings and Inches is highly significant by the Bradley and Schumann (1957) test ($w' = 7.94$, $a' = 4.23$, $b = 76$, $P < .05$, minimum $w'_{.05} = 3.18$).

Note the apparent inconsistency of the Phase 1 and Phase 3 findings. For Phase 1 both response languages show context effects in the same direction with Ratings being larger. For Phase 3 the context effects are in different directions, with Inches being greater. In the graphic presentation of trial-by-trial results of Fig. 1 and 2, the Inches judgments show complete reversal with the reversal of contexts, the lines crossing in the transition phase. In contrast, the Ratings never cross over, the direction of difference in Phase 3 remaining the same as in Phase 1.

Minor Findings.—In the Phase 3 analysis, the effect of Preshift Training is significant at the marginal 5% level for the rating groups. Those groups with shorter Phase 1's give lower ratings in Phase 3. This main effect is almost wholly contributed by the Low-High groups with the Short Phase 1.

The findings of the Phase 1-Phase 3 Shift analysis are not to be confused with a test of the significance of the Phase 1-Phase 3 shift, the uniformly high significance of which is reported below. Rather, the Phase 1-Phase 3 shift analysis shown in Table 2, examines the effects of treatments other than Phase upon the amount of shift. The score employed for each *S* is the difference in mean judgment between the last 4 trials of Phase 1 and the 10 trials of Phase 3, the direction of subtraction being reversed between High-Low and Low-High groups so that all changes in the expected contrast direction are positive. The analysis was designed primarily to demonstrate the expected effects of Discriminability and Preshift Training on stability of the adaptation level, or upon the differential occur-

TABLE 2

F VALUES FROM ANALYSES OF VARIANCE

Source	*df*	Phase 1 Judgments		Phase 3 Judgments		Phase 1-Phase 3 Shift	
		Inches	Ratings	Inches	Ratings	Inches	Ratings
Order (context) (O)	1	11.95**	254.79***	40.99***	5.16*	2.42	18.48***
Preshift training (T)	1	.79	2.27	.29	5.12*	1.99	4.16*
Discriminability (D)	1	1.40	1.34	.47	3.47	1.22	.22
O × T	1	.05	.02	.52	2.66	.54	1.85
O × D	1	.55	1.02	.01	.43	1.14	1.08
T × D	1	.54	.34	.23	.37	.97	.57
O × T × D	1	.00	.25	.20	.62	.02	.01
Within groups (*MS*)[a]	152	(23.20)	(60.35)	(21.52)	(84.27)	(8.72)	(42.98)
Inhomogeneity of variance	1/20	1.60	1.85	2.42	1.24	2.79	4.87*

[a] Mean squares shown in parentheses, rather than *F*s.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

rence of assimilation and contrast illusions among *S*s. Of the 320 *S*s, only 20 had negative scores and these were individually insignificantly different from no shift at all. Thus, this setting produced no persons who were significant assimilators, unlike two other settings in which these effects have been examined (Campbell, Hunt, & Lewis, 1957, 1958; Campbell et al., 1958).

The significant main effect for Order for shift scores is conceptually an interaction with the counterbalancing treatments. The shift is greater under the Low-High order than under the High-Low. This is very significant for Ratings and the trends for Inches are in the same direction. Note that the significant heterogeneity of variance casts suspicion upon this one analysis.

For Discriminability treatments, all effects are convincingly absent. For Preshift Training, effects are generally absent, with the 5% level findings for Ratings being in the direction of greater shift under Short Phase 1 treatments, a finding similar to that of Johnson (1949). While this effect is primarily contributed by the one Low-High Short Rating group, which also accounts for the Phase 3 Training main effect, slight trends are in the same direction for High-Low Ratings, and also for the two Inches comparisons.

Possible Interpretations.—To return to the initial findings presented, our preferred interpretation for the striking discrepancies between the Phase 1 and Phase 3 analyses is this: The symmetrical and reversible findings for Inches represent the operation of a single perceptual contrast process. This process is also present, and reversible, for the Ratings judgments, but in addition, the Phase 1 Ratings data contain a second process of defining perceptual equivalents for novel judgmental terms. That this operates in an adaptation level (Helson, 1947, 1948, 1959) fashion is indicated in Table 1 by the closeness with which the mean ratings of *all* stimuli approach the designated "average" value of 100 in all treatments. Of the Phase 1 effects of context on Ratings (of the tracer Stimulus 7) about one-half is attributed to this adaptation-level definitional process, the other half to the perceptual contrast effect. This definitional process is deemed irreversible without explicit retraining involving instructional feedback. This irreversibility is particularly strong under our instructions, "always use the same number for the same length of line," but would probably be so under usual single stimuli instructions, except in so far as "vocabulary exhaustion" (Campbell, Lewis, & Hunt, 1958) forced redefinition. In Phase 3, the experimentally reversed perceptual contrast effect and the perseverating definitional contrast effects tend to cancel each other just as they augmented each other in Phase 1.

This interpretation leads to the expectation that the magnitude of the Phase 1-Phase 3 shift for both Inches and Ratings should be the same, inasmuch as both are primarily affected by perceptual contrast effect, resulting from the experimental context reversal. While our choice of a common graphic scale for Fig. 1 and 2 was a somewhat unsatisfactory compromise, the method of selecting it was independent of the Phase 1-Phase 3 shifts. Hence, the degree to which the shift magnitudes turn out to be similar for the two judgmental languages encourages this interpretation. Note particularly in Fig. 1 the parallel course which the two High-Low lines take, and similarly the two Low-High lines. Graphic superposition (not shown here) shows them to be practically indistinguishable. We anticipate that were a more unequivocable psychophysical scale equivalence to be achieved, as through a study in which all 13 stimuli occur in each "trial" with an unchanging "phase," the magnitudes of our Phase 1-Phase 3 shifts would not significantly differ when translated into the common metric.

Some other indirect evidence for the equivalence of shifts under the two conditions comes from the t ratios for shift, computed by pooling all Discriminability and Preshift Training conditions. For Ratings and Inches, respectively, these were 11.92 and 11.33 for Low-High, and 10.95 and 11.34 for High-Low ($df = 79$, $P < .001$, each case). If the t ratios can be used as comparable measures of magnitude of effects (since the number of *S*s are equal), the effects are strikingly equivalent, with not the slightest trend toward statistical significance of difference. This as evidence for equivalence is of course quite inferential. However, the mode of inference is quite independent of that based upon the conversion formula employed in the graphs. The approximate agreement of these two independent forms of inference is taken as encouraging the retention of the hypothesized two-process interpretation.

The temporal weighting of experience.—The attractiveness of the two-process interpretation may be increased by considering an alternative uniprocess theory which would presume quite different weighting parameters in the temporal accumulation of experience under the two conditions. Three temporal weighting outcomes can be considered: a primacy effect, in which the first visual experiences would have greater effect than later ones in determining the adaptation level at any given time; a recency effect, in which the adaption level would be predominantly determined by very recent experience; and an equal temporal weighting outcome, in which all experiences (within the experimental session at least) would be equally weighted. Inches judgments, as seen in Fig. 1 and 2, clearly show a strong recency effect. The reversibility is almost complete, the two curves cross almost at the center of the transition phase. Apparently only the current trial stimuli or the ones just preceding are having much influence on the adaptation level against which the tracer stimulus is being judged.

In contrast, the Ratings fit perfectly a model of equal temporal weighting. Under this assumption, the judgments of Stimulus 7 for the High-Low and the Low-High groups should converge at that trial at which the total of experimental stimuli presented was the same, for at that point, each would be judged against the same adaptation level. For the Short treatment, this would be at Trial 14, i.e., when the number of Phase 1 and Phase 3 trials were equal. This is exactly what happens. For the Long groups, this would happen at a projected Trial 33. From Fig. 1 it seems plausible that this might occur although a slight primacy effect is also plausible.

While each of these uniprocess interpretations is plausible, it does not seem reasonable that so small a difference in conditions as the reporting language could make such a fundamental difference in the perceptual and memory processes as to provide an equal weighting in one instance and a recency weighting in another. Furthermore, a time-specific recency effect is probably the general rule for all nervous system processes, both end-organ adaptation and memory. For example, end-organ adaptation as involved in judgments of temperature, salinity, brightness, etc. all show recency effects so dominant that the primacy-recency problem, or the temporal weighting problem in any form, is almost never raised. In judgment studies, Parducci (1954, 1959) and Campbell, Lewis, and Hunt (1958) have found recency effects, and recency effects are to be inferred from Johnson's (1949) data. Recency effects likewise seem universal in learning, even if concomitant with and occasionally disguised by primacy effects (e.g., Miller & Campbell, 1959).

Accepting the dual-process interpretation would lead to the conclusion that inferences as to temporal weighting made from single-reversal studies using the method of single stimuli in its usual forms are erroneous. These would include Helson's early experiment (1947, pp. 16–17) used to justify equal temporal weighting, and the studies of Campbell, Hunt, and Lewis (1957, 1958) and Segall (1959). The two-process interpretation would predict that recency weighting is more general and that apparent equal weighting or primacy effects are a function of the semantic rather than perceptual effects. The evidence assembled here is not crucial for the two-process interpretation. A more definitive approach would be multiple reversal studies of a Low-High-Low-High nature. With increasing numbers of reversals of context, the semantic factors should decrease in salience, allowing the perceptual aspect to be manifest more directly.

SUMMARY

The effects of divergent and shifting stimulus contexts upon the judgment of a common tracer stimulus were measured for the task of judging lengths of lines projected on a screen. For judgments made in inches, a reversible perceptual-contrast effect was demonstrated. Judgments made with an uncrowded numerical rating scale showed in addition a response-definitional contrast effect augmenting the apparent contrast effect in the initial phase and persisting into subsequent phases. These findings cast doubt upon previous estimations of the appropriate temporal weighting of stimuli in contributing to an adaptation level. Effects of stimulus discriminability and preshift training are also examined.

REFERENCES

BRADLEY, R. A., & SCHUMANN, D. E. The comparison of the sensitivities of similar experiments: Theory. *Ann. math. Statist.*, 1957, **28**, 902–920.

CAMPBELL, D. T. Factors relevant to the validity of experiments in social settings. *Psychol. Bull.*, 1957, **54**, 297–312.

CAMPBELL, D. T., HUNT, W. A., & LEWIS, N. A. The effect of assimilation and contrast in judgment of clinical materials. *Amer. J. Psychol.*, 1957, **70**, 347–360.

CAMPBELL, D. T., HUNT, W. A., & LEWIS, N. A. The relative susceptibility of two rating scales to disturbances resulting from shifts in stimulus context. *J. appl. Psychol.*, 1958, **42**, 213–217.

CAMPBELL, D. T., LEWIS, N. A., & HUNT, W. A. Context effects with judgment language that is absolute, extensive, and extra-experimentally anchored. *J. exp. Psychol.*, 1958, **55**, 220–228.

HELSON, H. Adaptation-level as a frame of reference for the prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 297–313.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 1–29.

HELSON, H. Adaptation-level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

JOHNSON, D. M. Learning function for a change in scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851–860.

KRANTZ, D. L. Stimulus and response relationships in absolute judgment. Unpublished master's thesis, Northwestern University, 1959.

MILLER, N., & CAMPBELL, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *J. abnorm. soc. Psychol.*, 1959, **59**, 1–9.

PARDUCCI, A. Learning variables in the judgment of single stimuli. *J. exp. Psychol.*, 1954, **48**, 24–30.

PARDUCCI, A. An adaptation-level analysis of ordinal effects in judgment. *J. exp. Psychol.*, 1959, **58**, 239–246.

SEGALL, M. H. The effect of attitude and experience on judgment of controversial statements. *J. abnorm. soc. Psychol.*, 1959, **58**, 61–69.

STEVENS, S. S. Adaptation level vs. the relativity of judgment. *Amer. J. Psychol.*, 1958, **71**, 633–647.

(Received April 30, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 43-50

EFFECTS OF FOREPERIOD, FOREPERIOD VARIABILITY, AND PROBABILITY OF STIMULUS OCCURRENCE ON SIMPLE REACTION TIME[1]

D. H. DRAZIN[2]

Oxford University

Only recently has any systematic attempt been made to elucidate the relationship between the simple reaction time (RT), the foreperiod of the reaction (interval between warning signal and stimulus), and the variability of the foreperiod. Klemmer (1956) found that when the foreperiod was varied at random within a rectangular frequency distribution, mean RT increased as the mean foreperiod and range of foreperiods were separately increased. In a further study, employing normal distributions of foreperiods, Klemmer (1957) adduced evidence in support of the hypothesis that the mean RT varies in direct proportion to *S*'s uncertainty as to the timing of the stimulus, this uncertainty arising from two sources, variable error in *S*'s time-keeping and objective foreperiod variability. He estimated the former component on the basis of "prediction" tests, in which *S* was required to reproduce intervals equal to the mean foreperiods of matched simple reaction tasks. Corresponding foreperiod and prediction variances were summed and a conversion made to information measure (bits) relative to the uncertainty of the individual *S*'s predictions of a 1-sec. interval.

Although Klemmer's "informational hypothesis" offers an explanation of the relationship between mean RT and the parameters of the foreperiod frequency distribution, it provides no account of the relationship between RT and the foreperiod of the individual reaction (immediate foreperiod). While it must be conceded that Klemmer (1956) found no evidence of such an effect, contrary findings (Breitweiser, 1911; Karlin, 1959; Mowrer, 1940; Telford, 1931; Woodrow, 1914) of some consistency are on record and evidently require some explanation.

The present experiments, employing continuous rectangular foreperiod distributions, gave rise to pronounced RT-foreperiod curves comparable to those previously reported. It is shown that the form of the RT-foreperiod curve is systematically dependent on the parameters of the foreperiod frequency distribution. The effect of probability of stimulus occurrence was also examined. This variable is of special interest in connection with the common practice of administering a small proportion of check trials (wherein the stimulus does not occur) so as to minimize the frequency of anticipatory reactions.

METHOD

Apparatus.—The *S* received a 50-cps warning tone, which also served to mask extraneous noise in and about the laboratory, through close-fitting headphones. The stimulus was the onset of a Hivac CCIL neon lamp which *S* observed against a black background at a distance of 18 in. The warning tone and stimulus remained on until *S* had responded by pressing a reaction key with the preferred index finger. The time recording equipment was stopped when a force of 120 gm. was applied to the key.

---

[1] Based on a thesis submitted in partial fulfilment of the requirements for the DPhil degree at Oxford University. The author wishes to express his gratitude to his supervisor, H. Kay, for advice and encouragement, and to the Department of Scientific and Industrial Research for financial support.

[2] Now at the RAF Institute of Aviation Medicine, Farnborough, England.

Foreperiod intervals were produced at random within a continuous rectangular distribution, for preset values of minimum foreperiod and range of foreperiods, by a system of relays and condenser-resistance time-delays. One time-delay, preset to the required range of foreperiods, cycled continuously, closing a relay momentarily at regular intervals. Another time-delay, initiated by *E*, closed a self-latching relay after an interval preset to the required minimum foreperiod. The simultaneous closure of both relays triggered a final self-latching relay which switched on the stimulus lamp. In order to ensure that *E* started the sequence of events in random phase with the cycling timer, all relays were silenced and care was taken to render the variability of the intertrial interval large in relation to that of the longest range of foreperiods employed. These precautions succeeded in producing foreperiod distributions which were not to inspection discriminable from rectangular.

The occurrence or nonoccurrence of the stimulus on a given trial was determined by the terminal state of a uniselector which rotated for a random period (the foreperiod) and which was so connected that a circuit between a selected set of wipers and the central pole of the selector switch was made on random trials according to a preset probability.

Individual foreperiods and RTs were recorded as cycle counts of a 1000 cps quartz crystal oscillator on decade counters.

TABLE 1

DESCRIPTION OF CONDITIONS: EXP. I AND II

Experiment I[a]			Experiment II[b]	
Cond.	Range of Foreperiods (Sec.)	Probability of Stimulus Occurrence	Cond.	Minimum Foreperiod (Sec.)
A	2.0	1.0	P	2.0
B	2.0	.9	Q	1.0
C	2.0	.5	R	0.5
D	1.0	1.0	S	0.25
E	1.0	.9	T	0.125
F	1.0	.5		
G	0.5	1.0		
H	0.5	.9		
J	0.5	.5		

[a] Mean foreperiod = 1.5 sec.
[b] Range of foreperiods = 1.0 sec.; probability of stimulus occurrence = 1.0.

Procedure.—Two separate experiments are reported. The conditions are described in Table 1. In Exp. I, mean foreperiod was held constant throughout at 1.5 sec., range of foreperiods and probability of stimulus occurrence serving as variables in a 3 × 3 factorial design.

The *Ss* attended 5 practice and 12 test sessions, each of 50 min., over a period of three weeks. The practice sessions, comprising 30 trials in each condition, were run daily during the first week, the conditions being held in the same order every day.

The test sessions were run daily during the following two weeks. All probabilities of stimulus occurrence were always given in immediate succession for each range of foreperiods; though sequential effects of range, and of probability within each level of range, were completely counterbalanced within each set of six sessions. Ten warm-up trials were held prior to the first test run at each range, every run consisting of 30 trials as in the practice sessions.

In all sessions *S* was informed of the conditions of each ensuing run. He was instructed to respond as soon as possible after seeing the stimulus, while taking care not to respond in anticipation. In the practice periods, he was cautioned after responding prematurely. He was informed of his RT after each reaction and of his mean score in every condition after each session. A rest interval of 2 min. was allowed between runs. The experiment was held in a dimly lit room. The mean intertrial interval was about 10 sec.

In Exp. II, range of foreperiods was held constant at 1.0 sec. and probability of stimulus occurrence at 1.0, minimum foreperiod being held at five levels differing by factors of two. The *S*s attended 5 practice and 10 test sessions, each of 30 minutes, over a period of three weeks. In the practice sessions, conditions were given in the same order every day. In the experiment proper, order effects were counterbalanced within each set of five sessions. Each run consisted of 50 trials.

Subjects.—The *S*s were three University students aged between 19 and 26. Two *S*s served in each experiment.

RESULTS[3]

Experiment I.—The data were treated as follows. Taking *S*s and

---

[3] Complete data, consisting of mean RTs, *SD*s, and *N*s for all points in Fig. 1 and 2, unweighted mean RTs, and frequencies of anticipatory reactions, have been deposited with the American Documentation Institute.

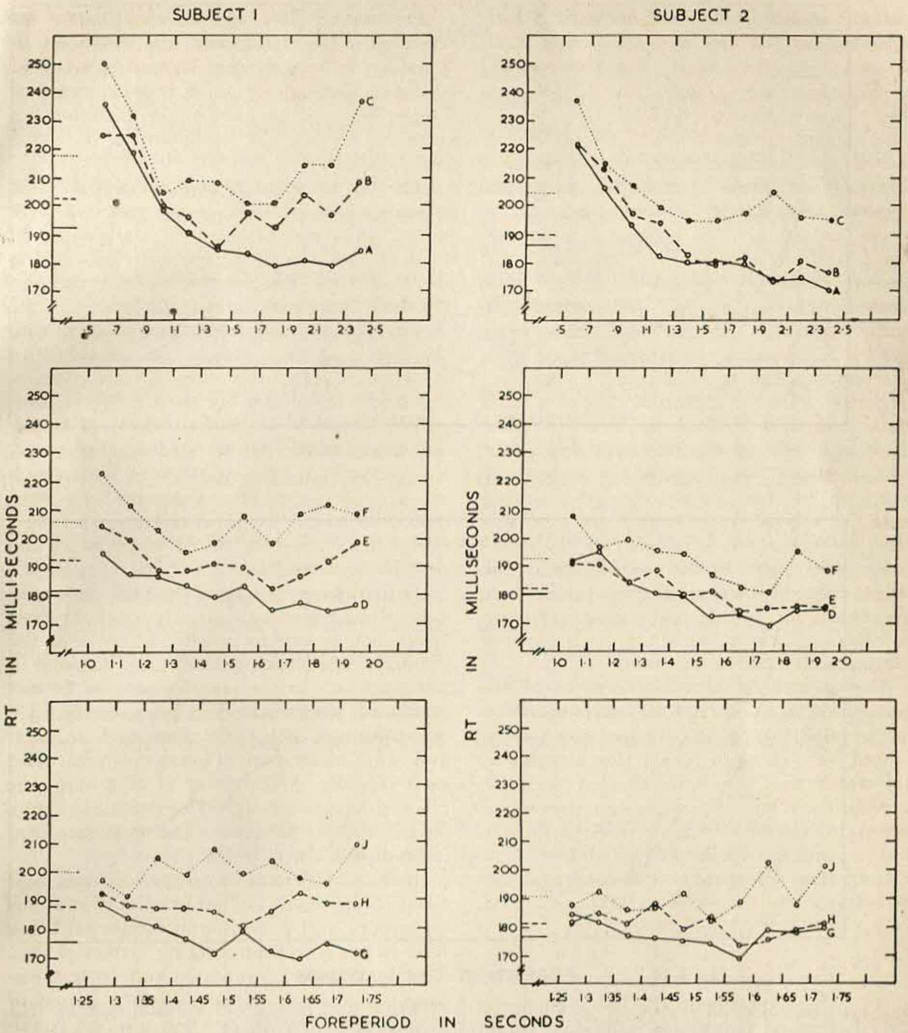


FIG. 1. Experiment I: The influence of foreperiod, foreperiod variability, and probability of stimulus occurrence upon RT. (RT-foreperiod curves for individual conditions and *S*s—see Table 1. Overall unweighted mean RTs indicated on the ordinate by horizontal straight lines.)

conditions separately, the range of foreperiods was divided into 10 equal sections and the mean of RTs associated through their foreperiods with each section was plotted against the expected section mean. The resultant RT-foreperiod curves are illustrated in Fig. 1, the horizontal lines on the ordinate indicating the overall mean RTs. False reactions (occurring before the stimulus) and premature reactions (occurring after, but presumed to have been initiated in the central nervous system before, the

---

Order No. 6683 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

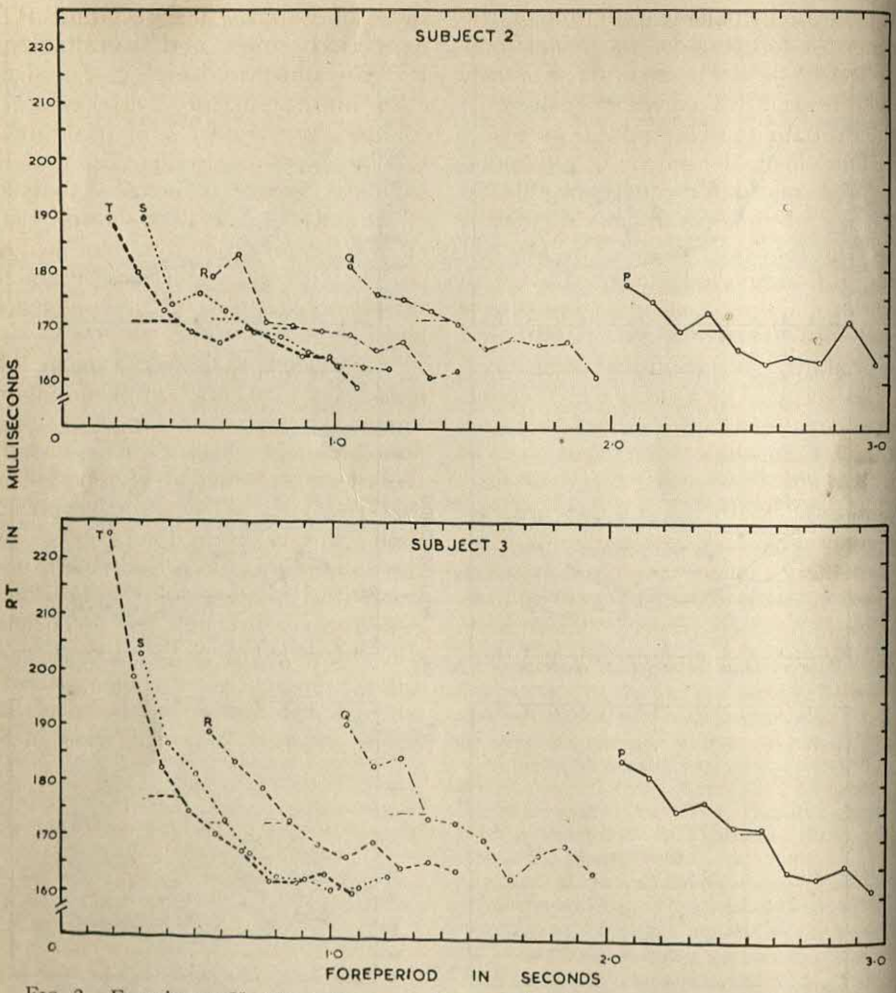


FIG. 2. Experiment II: The influence of foreperiod and foreperiod variability upon RT. (RT-foreperiod curves for individual conditions and *S*s—see Table 1. Overall unweighted mean RTs indicated on regression curves by horizontal straight lines.)

stimulus) were excluded from this treatment. A reaction was discarded as premature if *S* judged it to be so or if the recorded RT was less than 100 msec. Since false and premature responses accounted for only 1.5% of trials, it is concluded that anticipation was a negligible source of error.

Analysis of variance of raw RTs was performed in order to determine whether the illustrated RT-foreperiod curves were compatible, at the 2.5% confidence level, with the hypothesis of simple regression.[4] Significant deviations from simple regression were obtained in all conditions except E, H, and J (S_1) and G, H, and J (S_2).

It may be seen that overall mean

---

[4] Slight heterogeneity of variance was noted but, in view of the different and complex relationships which appeared to hold between the cell means and variances in the several conditions of the experiment, it was decided not to attempt a transformation. Allowing for this source of error, the results of the analyses may safely be regarded as significant at the 5% level.

RT is an increasing function of the range of foreperiods, the mean foreperiod being held constant. Further, overall mean RT varies inversely with the probability of stimulus occurrence.

The effect of immediate foreperiod is pronounced for ranges of 1.0 and 2.0 sec., but generally nonsignificant for the 0.5-sec. range. The form of the RT-foreperiod curve varies with the probability of stimulus occurrence and also between *S*s. With unit probability of stimulus occurrence, RT decreases as a negatively accelerated function of immediate foreperiod, this relationship being most marked in Cond. A (2.0-sec. range of foreperiods). With the 2.0-sec. range of foreperiods, S_1's results show an increasingly steep positive terminal gradient with decreasing probabilities of stimulus occurrence. There is no indication of such a gradient in the results of S_2.

Experiment II.—The data of Exp. II were treated in the same way as those of Exp I. The resultant RT-foreperiod curves and overall mean RTs are illustrated in Fig. 2. Since false and premature responses accounted for only 1.2% of trials, it is concluded that anticipation was a negligible source of error. Analysis of variance of raw RTs showed that all RT-foreperiod curves were incompatible with the hypothesis of simple regression at the 0.1% confidence level.

It is evident that overall mean RT is not a critical function of minimum foreperiod when the range of foreperiods is held constant. As in Exp. I, the effect of immediate foreperiod is marked, long RTs again occurring with relatively short foreperiods. As the minimum foreperiod increases, the initial gradient of the RT-foreperiod curve becomes less steep and the curve as a whole more linear. Inter-*S* differences are again apparent: whereas the forms of the two *S*s' curves are broadly similar, those of S_3

TABLE 2

Simple and Interaction Effects (in Msec.) of Immediate Foreperiod (Fp), Preceding Foreperiod (Fp$_{-1}$), and Second Preceding Foreperiod (Fp$_{-2}$)

Effect	Condition				
	P	Q	R	S	T
Subject 2					
Fp	−6.82**	−9.37**	−9.44**	−12.19**	−10.12**
Fp$_{-1}$	−0.01	4.87**	3.47*	2.99*	1.18
Fp$_{-2}$	−0.53	1.11	−1.28	3.45*	0.17
Fp × Fp$_{-1}$	0.13	−1.91	−1.25	−1.07	1.85
Fp × Fp$_{-2}$	−1.21	1.69	0.37	2.54	1.17
Fp$_{-1}$ × Fp$_{-2}$	−0.09	−2.61	0.63	1.01	−0.76
Fp × Fp$_{-1}$ × Fp$_{-2}$	−1.44	−0.24	1.97	−1.13	−1.66
Estimated *SE*	1.35	1.33	1.64	1.37	1.52
Subject 3					
Fp	−13.18**	−13.93**	−13.83**	−20.96**	−27.76**
Fp$_{-1}$	4.14*	5.32*	7.40**	5.45**	0.60
Fp$_{-2}$	−0.68	1.96	3.52*	4.00*	2.03
Fp × Fp$_{-1}$	−1.21	−5.86**	−4.31*	−0.44	0.58
Fp × Fp$_{-2}$	−0.00	2.00	−0.01	−3.02	1.06
Fp$_{-1}$ × Fp$_{-2}$	−4.54*	−2.29	1.53	2.93	−0.74
Fp × Fp$_{-1}$ × Fp$_{-2}$	0.33	−0.05	0.09	−0.41	2.72
Estimated *SE*	1.75	1.72	1.39	1.66	1.97

* $P < .05$.
** $P < .01$.

show a more marked foreperiod effect.

The *SD*s about values plotted in Fig. 2 appear to vary systematically with the length of foreperiod relative to the range of foreperiods (see supplementary data). The results of both *S*s exhibit a similar trend in each condition: the scatter of RTs is greatest at relatively short intervals, decreases steeply to a minimum at about the mean interval, and then increases gradually towards the longest interval.

Previous workers (Karlin, 1959; Klemmer, 1956; Woodrow, 1914) have shown that the RT on any given trial is influenced to some extent by the lengths of preceding foreperiods relative to the mean foreperiod, longer RTs occurring when the foreperiod of the preceding trial is relatively long. Klemmer (1956) sought but failed to find any effects due to the foreperiod of the second preceding trial. The results of an analysis of the effects of the immediate foreperiod (Fp), the preceding foreperiod (Fp_{-1}), and the second preceding foreperiod (Fp_{-2}), based on the data of Exp. II, are presented in Table 2. Taking *S*s and conditions separately, all RTs, other than those recorded on the first and second trials of a run, were entered in a $2 \times 2 \times 2$ classification according to whether Fp, Fp_{-1}, and Fp_{-2} were separately greater or less than the expected mean foreperiod. The magnitudes of all main and interaction effects were estimated by taking sums and differences of means in the usual way. Significance was determined on the basis of estimated standard errors.

The fact that the main effects of Fp are all negative and highly significant indicates that short immediate foreperiods gave rise to long RTs, merely confirming the finding established by a more precise analysis of variance that the RT-foreperiod curves depart significantly from simple regression.

There would appear to be a discontinuity in the influence of preceding foreperiods between minimum foreperiods of 0.125 and 0.25 sec. In Cond. T (minimum foreperiod = 0.125 sec.) only the effect of Fp is significant, whereas several effects are significant in other conditions.

Discounting Cond. T, long RTs tend to follow trials with short foreperiods and vice versa, as shown by the fact that the main effects of Fp_{-1} are mostly positive and significant. There is a suggestion that Fp_{-2} has an influence similar to but weaker than that of Fp_{-1}. This suggestion is confirmed for S_3, the mean effect of Fp_{-2}, averaged over Cond. P, Q, R, and S, exceeding the compound *SE* by a factor of 2.70 ($P < .01$).

That the $Fp \times Fp_{-1}$ interactions are mostly negative suggests that the influence of the immediate foreperiod is greatest following trials with short foreperiods. This is confirmed for S_3, the mean value of this effect, averaged over Cond. P, Q, R, and S, exceeding the compound *SE* by a factor of 3.62 ($P < .001$).

It may also be noted that the effects of preceding foreperiods tend to diminish as the minimum foreperiod increases beyond about 0.5 sec. The $Fp \times Fp_{-2}$, $Fp_{-1} \times Fp_{-2}$, and $Fp \times Fp_{-1} \times Fp_{-2}$ interactions are generally nonsignificant.

DISCUSSION

The results of the present study tend to confirm Klemmer's (1956) finding that overall mean RT, irrespective of immediate foreperiod, is an increasing function of the range of foreperiods and that where the range is appreciable (say more than 0.5 sec.) overall mean RT is not a critical function of the mean foreperiod. It is notable that S_2's and S_3's overall mean RTs in the five conditions of Exp. II are narrowly scattered over ranges of about 2 and 5 msec., respectively. Similarly Klemmer, using a 2.0-sec. range, obtained the same overall mean RT with minimum foreperiods of 0.25 and 3.25 sec.; with a minimum foreperiod of 6.25 sec. the overall mean RT was only about 13 msec. longer.

Klemmer's (1957) hypothesis that overall mean RT is a linear function of *S*'s uncertainty as to the timing of the stimulus is contraindicated by the present data. Woodrow (1930) found that, for intervals less than about 4 sec., the variability of temporal judgments (reproduction of temporal intervals) was roughly proportional to the time judged, the *SD* being about 10% of the standard interval. Thus, according to Klemmer, we should expect that, when range is held constant, overall mean RT would be an increasing function of the minimum foreperiod. It has been noted, however, that the results of Exp. II showed no trend in this direction.

Most of the RT-foreperiod curves obtained in the present study and many illustrated in the literature (e.g., Breitweiser, 1911; Karlin, 1959; Woodrow, 1914) exhibit marked initial gradients. Neither the characteristics nor even the existence of such gradients are explicable in terms of Klemmer's "informational hypothesis," since this only relates time uncertainty to the overall mean RT. Given the finding that RT is related to the immediate foreperiod, then it would seem that the characteristics of this relationship must receive prior explanation: a theory which could account for the entire RT-foreperiod curve would a fortiori account for the overall mean RT. That Klemmer failed to obtain evidence of an initial gradient in the RT-foreperiod relationship may possibly be due to a fundamental difference in experimental procedure. Klemmer used ranges composed of 11 equally spaced foreperiods delivered in random order (i.e., a discrete foreperiod frequency distribution bound by a rectangular envelope). For very short intervals, where the difference between alternative foreperiods (interforeperiod difference) must have been easily discriminable, *S* may have learned to "focus his attention" at or about the moments when there was a definite probability that the stimulus would occur, and by so doing he might have been able to respond more quickly (Mowrer, 1940). In the present study, however, foreperiod variability was continuous, so that such tactics would have been inappropriate. Concordant with this suggestion is the fact that in Klemmer's (1956, Fig. 1) variable foreperiod tests, where the interforeperiod difference was of the same order of magnitude as the minimum foreperiod, the mean RTs for the shortest intervals were below the overall mean RTs for their respective ranges but, where the interforeperiod difference was relatively small, the mean RTs for the first intervals were, as in the present study, somewhat above the general means.

Woodrow (1914) and Karlin (1959) have suggested that the RT-foreperiod relationship reflects progressive changes in *S*'s state of readiness. According to this view, the effects of preceding foreperiods are attributable to an adaptation process: on average, readiness reaches a peak or plateau at about the mean interval, but this peak or plateau tends to be attained earlier or later, according to the length of the preceding interval. The present results confirm the findings of Karlin (1959), Klemmer (1956), and Woodrow (1914) that RT varies as an increasing function of the preceding foreperiod. It has also been shown that RT varies in a like manner but to a less extent with the second preceding foreperiod. This suggests that the adaptation process which has been postulated is heavily biased in favor of recent intervals, but may be sensitive to quite remote ones.

The present finding that the scatter of RTs is greatest at relatively short intervals and least at intermediate ones may be interpreted as further support for the adaptation hypothesis. Readiness will be most variable at the minimum foreperiod, since on some occasions, following a series of relatively long intervals, the stimulus will occur at a point of minimum readiness, whereas on other occasions, following a series of relatively short intervals, the stimulus will occur at a point of maximum readiness. At intermediate intervals, however, *S*'s state of readiness will rarely be much less than maximal and a smaller scatter of RTs might therefore be predicted.

It is interesting to note two areas in which the evidence concerning the effects of preceding foreperiods is in apparent conflict. First, while Karlin (1959) found that the effect of the immediate foreperiod was greatest following long preceding foreperiods, the present results show this effect to have been greatest following short preceding foreperiods. Further, it would seem that Klemmer (1956) found no significant trend in either direction.

Second, whereas Karlin obtained anomalous but significant effects due to preceding foreperiods in conditions where the minimum foreperiod was less than 0.8 sec., the present results indicate no significant effects of previous foreperiods with very short minimum foreperiods.

It is to be hoped that discrepancies such as these will be resolved by further studies, so that the characteristics of the postulated adaptation mechanism can be specified more clearly. It might then be possible to predict the form of RT-foreperiod curve which would result from the use of any random distribution of foreperiods. The mean RT for any foreperiod could be calculated simply by weighting the RTs characteristic of all possible patterns of preceding foreperiods by their expected frequencies within the distribution in question.

In Exp. I it was found that the RT-foreperiod relationship was dependent on the probability of stimulus occurrence, and that the form of this dependence was subject to sizable individual differences. It would be premature to speculate on the general significance of these results or their specific relevance to the adaptation hypothesis discussed above. At present it will suffice to indicate the necessity of controlling the frequency of check trials in investigations into the form of the RT-foreperiod relationship.

SUMMARY

In two parallel experiments, the foreperiod of a simple reaction task was varied at random within a rectangular frequency distribution. In a total of 14 conditions, the minimum foreperiod, range of foreperiods, and probability of stimulus occurrence were systematically varied.

The results indicate that the RT-foreperiod relationship is subject to a range effect: in all conditions where the range of foreperiods exceeded 0.5 sec., RT tended to decrease initially as a negatively accelerated function of the foreperiod. The detailed form of this relationship was found to depend upon the minimum foreperiod, the range of foreperiods, and the probability of stimulus occurrence. The scatter of RTs was greatest for short foreperiods and least for foreperiods of intermediate length.

Reaction time was also found to vary with the foreperiods of the preceding reaction and, to a less extent, the second preceding reaction: in either case, long RTs tended to follow reactions with short foreperiods and vice versa. The effect of immediate foreperiod was most marked following reactions with short foreperiods.

The experimental findings are interpreted as evidence in favor of the view that RT reflects *S*'s state of readiness for the stimulus.

REFERENCES

BREITWIESER, J. V. Attention and movement in reaction time. *Arch. Psychol., NY*, 1911, **2**, No. 18.

KARLIN, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58**, 185–191.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

KLEMMER, E. T. Simple reaction time as a function of time uncertainty. *J. exp. Psychol.*, 1957, **54**, 195–200.

MOWRER, O. H. Preparatory set (expectancy): Some methods of measurement. *Psychol. Monogr.*, 1940, **52**(2, Whole No. 233).

TELFORD, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1–36.

WOODROW, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17**(5, Whole No. 76).

WOODROW, H. The reproduction of temporal intervals. *J. exp. Psychol.*, 1930, **13**, 473–499.

(Received May 3, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 51–57

FREE RECALL AND ORDERING OF TRIGRAMS[1]

LEONARD M. HOROWITZ[2]

Johns Hopkins University

Some explanation is needed for the fact that intralist similarity can have either inhibitory or facilitative effects on the learning and retention of verbal material. The specific role of intralist similarity seems to depend upon the nature of *S*'s task, although the critical task variables are not known. The purpose of the present study is to examine some of these critical variables.

Underwood (1952, 1954) varied the similarity relationship between the items of a list learned by serial anticipation and concluded that "recall and intralist similarity are not systematically related" (1954, p. 160). To vary intralist similarity, Underwood used lists of consonant-vowel-consonant (CVC) trigrams which contained as many different letters as possible (low intralist similarity) and other lists in which certain letters occurred repeatedly throughout the syllables (high intralist similarity). More trials were needed to learn, and later to relearn, the high similarity lists to criterion. But the number of overt errors per trial during learning was not related to intralist similarity, nor was recall 24 hr. later.

Underwood defines intralist similarity as the extent to which the items of a list share letters in common with one another. Intralist similarity can also be expressed in terms of the sequential restrictions between the letters of the items. For example, particular groupings of two or more letters might occur with a high probability in the syllables of one list while these relationships could be left to chance in another list. The list in which the sequential dependencies depart markedly from randomness would be called highly redundant (Miller, 1958).

Miller (1958) set up explicit rules to generate redundant lists of items four to seven letters long. He also used lists in which the item letters were combined at random. When redundancy was introduced into Miller's lists, the items became more similar (by Underwood's definition) to one another. Miller has demonstrated that the redundant lists had greater intralist similarity by a task in which *S*s search for particular items from a list of redundant items and from a list of random items. The median search time was considerably higher for the redundant lists.

Each of Miller's lists was presented in various orders, and after each presentation *S*s were tested by the method of free recall. Recall of the redundant lists (high intralist similarity) was superior to that of the random lists, not only throughout original learning but also after the interpolation of another list. Since Underwood found that recall and intralist similarity are not systematically related, this outcome is rather surprising.

Aborn and Rubenstein (1952; Rubenstein & Aborn, 1954) have found essentially the same result as Miller with verbal material more like Underwood's. Whereas Miller's material contained four different consonants only, Aborn and Rubenstein used 16 CVC trigrams involving 19 different letters. The syllables were put together to form a list, but the lists ranged from a random assortment of trigrams to a highly redundant one ("highly organized," in their terms). As in Miller's data, the higher the intralist similarity, the greater the mean recall score.

The measure of learning in the Aborn and Rubenstein studies was more like Miller's method of free recall than Underwood's serial learning since, for the most part, the serial order of *S*'s recall did not affect his score. If a trigram was correctly placed within a grouping of four items (whether the grouping itself was recalled early or late), it was scored as a correct item. It was also scored correct it if merely followed another trigram appropriately in recall.

Thus the principal difference between Underwood's study of serial anticipation and the studies by Miller and Aborn and Rubenstein is not in the concept needed to describe the interitem relationship (similarity), but in the measures of learning used. The major difference between serial learning and free recall is that serial learning requires a unique ordering of the items, while free recall does not.

---

[1] This paper is based upon a dissertation submitted in partial fulfillment of the requirements for the PhD degree at Johns Hopkins University. It reports work done while the author was a Fulbright Fellow at University College, London, and a Social Science Research Council Fellow at Johns Hopkins University. The author is indebted to James Deese and A. R. Jonckheere for their enthusiastic advice and encouragement.

[2] Now at Stanford University.

The closest experiments to a study of the ordering process are discrimination studies like that of Fitts, Weinstein, Rappaport, Anderson, and Leonard (1956). Discrimination, like the ordering part of serial learning, places a premium on the uniqueness of items in the test material. Fitts et al. found that randomly generated geometrical figures can be recognized more rapidly than statistically constrained figures. In other words, redundancy is a hindrance to visual pattern recognition. It would not be surprising, then, to discover that redundancy is a hindrance to ordering as well.

The evidence suggests that high intralist similarity may facilitate free recall learning and hinder ordering. When ordering and free recall are combined in the serial anticipation method, serial learning could appear unrelated to intralist similarity. In view of this possibility, the first step in an analysis of intralist similarity would be to study the effects of such similarity upon free recall and ordering separately. That is the purpose of the present experiment.

METHOD

Materials

Two lists of 3-consonant (CCC) trigrams were used. One list contained fewer different letters than the other and therefore had greater intralist similarity. The high-similarity list was composed of four letters and is referred to as L4. It made use of F, S, V, and X to make a total of 12 items. These items are VXF, XVS, FSV, SFX, XFV, FXS, VSX, SVF, VFS, FVX, SXV, and XSF. Each of the four letters occurs three times as the initial letter of a syllable, three times as the middle letter, and three times as the final letter. A single letter occurs only once in an item.

This list contains half of the 24 possible three-letter combinations. Moreover, any one item in the group can be transformed into another item by applying the following rule: Replace any one letter of a three-letter combination by the fourth letter and interchange the remaining two letters of that item. For example, the first item of the list is VXF. If the F of this item is replaced by an S, and the V and X are interchanged, the new item is XVS, the second item on the list. Application of this rule yields precisely the above 12 items and no others. The rule does not generate the reversals of the items, though. (FXV is a reversal of VXF and does not occur.)

Each pair of letters occurs twice—once as the first two letters of a syllable and once as the last two letters of a syllable. In both cases the pair occurs in combination with the same third letter. (FX, for example, occurs in the item FXS as well as SFX.)

The other list is called L12 because it contains 12 different letters. The L12 items are: QSZ, KXW, MRV, WQJ, VKH, ZMQ, HWS, FVX, JZR, XHF, SJM, and RFK. This list has the following properties: (*a*) Each letter occurs once in each position; (*b*) a particular pair of letters occurs in only one item; and, of course, (*c*) all possible pairs or triplets of the letters do not appear in the list.

Methods of Presentation

Each list was presented to two groups of *S*s. One group was required to learn the items by the method of free recall. The other was required to learn the *order* of the items, but not the items themselves.

Free recall.—L4 was presented one trigram at a time, on 4×5 in. flash cards. Each trigram was shown for 4 sec. After the *S*s had seen the entire list, they were instructed to record on their answer sheets, in any order, as many of the items as they could remember; guessing was permitted. The answers were collected, and the items were presented again in a different order. This was done in order to minimize self-instruction for ordering among the *S*s. A second group of *S*s was presented with L12 in the same way.

Ordering.—L4 was presented to a group of *S*s in a standard order of presentation. The order was the order in which the items are listed above. After each presentation, *S*s were supplied with a packet of 12 $1 \times 1\frac{3}{4}$ in. slips of paper. Each slip of paper had on it one of the L4 items. The *S*s were instructed to re-arrange these 12 slips so that the order corresponded to the order in which the flash cards had been presented. (The order of the items in the packet was different on each of the 10 trials.)

A fourth group ordered the L12 items under the same procedure. The standard order of the L12 items was the order in which they are listed above.

TABLE 1

MEAN CORRECT RECALL SCORES AND MEAN TOTAL RESPONSE SCORES WITH L4 AND L12

Trial	Mean Number of Syllables Correctly Recalled		Mean Number of Responses Given	
	L4	L12	L4	L12
1	3.12	1.83	7.20	6.08
2	4.80	3.54	9.24	7.75
3	5.72	4.79	10.08	8.62
4	6.24	5.96	10.40	10.08
5	6.56	5.92	10.52	9.79
6	6.64	7.12	10.64	10.54
7	6.60	7.42	10.20	10.42
8	7.00	8.00	10.68	10.42
9	7.16	8.46	10.96	10.08
10	7.79	8.50	10.44	11.08

Subjects

Each group contained the children of a preformed class in the Sir William Collins School, a secondary modern school in London, England. These groups were comparable in age, IQ, and academic achievement, and *S*s within each group were tested simultaneously. There were 24 and 25 *S*s in the free recall groups, and 30 and 19 *S*s in the ordering groups. (The groups were unequal in size since the classes were unequal in size, but this difference does not affect the statistical analysis of the data.) All *S*s were male, 14 or 15 yr. old, and slightly above average in intelligence.

RESULTS AND DISCUSSION

The most general result is that free recall for L4 is higher than for L12 early in learning (although the superiority is lost on later trials), while L12 is consistently easier to order than L4. The detailed results for the two methods are reported below.

Free Recall

Correct responses.—Table 1 shows the mean number of syllables correctly recalled on each trial and the mean number of responses (correct and incorrect) that *S*s produced. L4 is superior to L12 in correct recall for Trials 1 through 5. Tested by the Mann-Whitney *U* test, the initial difference is significant at $P < .001$. Table 2 summarizes the analysis of variance of the correct recall scores. The Trial × List interaction is significant at $P < .001$, although on the average, the difference between the two lists is not significant.

The superiority of L4 in free recall is probably the result of chance guessing. Each L4 *S* must have learned all the L4 letters quite early since each *S* used all four letters in his responses on every trial. Furthermore, each response was a three-letter combination. If *S* randomly combined the four L4 letters into three-letter combinations, the probability of his producing a correct item is $\frac{1}{2}$. Table 1 shows that the mean correct score for L4 on the early trials is about half the mean total response score. Thus, the advantage of the high-similarity list is that it enables *S*s to correctly guess a large number of items.

Just why L4 lost its superiority, though, is not completely clear. This loss of superiority may be due in part to the extreme similarity of the L4 items to each other. In the later stages of free recall learning, two or more very similar items may general-

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT RECALL SCORES

Source of Variance	*df*	*MS*	*F*
Between *S*s	48		
Lists	1	0.012	—
Error (b)	47	21.365	
Within *S*s	441		
Trials	9	151.269	50.79*
Lists × Trials	9	11.797	3.96*
Error (w)	423	2.978	
Total	489		

* $P < .001$.

ize to the same erroneous response at the expense of a correct item. Such generalization may be difficult to overcome so that further improvement is hindered. This issue needs to be studied further with more levels of intralist similarity.

Intrusions.—There were 198 letter-intrusions throughout the responses of L4. E, U, and Y made up 72% of these intrusions. The responses of L12 contained 140 intrusions, and all 14 possible intrusions occurred. Since *S*'s recall was handwritten, the absolute number of intrusions may be slightly biased by possible errors in scoring. Therefore, a statistic summarizing the variability of intrusions is probably more appropriate here: the uncertainty in the L12 intrusions was 3.36 bits, while that in the L4 intrusions was 2.14 bits. Thus, L12 led to a greater variety of intrusions than L4. One reason for this difference in distribution is that the 12 letters of L12 evoked more different verbal habits than did the 4 letters of L4.

Ordering

Kendall's *P* (1955, pp. 3–7) was selected as the best measure of "goodness of ordering." All pairs of items are considered one by one in *S*'s returned packet; and whenever two items are correctly placed relative to one another, +1 is scored for that pair. Kendall's *P* may be considered the total number of +1's in a particular ordering.[3] $P = 66$ and $P = 0$ indicate perfect ordering and perfect reverse ordering, respectively, of 12 items. $P = 33$ is the mean chance expectation and indicates no relationship between the correct order and *S*'s order. The *SD* of *P* values for chance orderings of 12 items is 7.3.

TABLE 3

MEAN ORDERING SCORES AND MEAN CORRECT PLACEMENT SCORES WITH L4 AND L12

Trial	Mean Ordering Score (Kendall's *P*)		Mean Correct Placements	
	L4	L12	L4	L12
1	35.48	52.33	1.44	4.70
2	41.21	55.14	3.41	5.79
3	43.70	59.07	2.78	7.73
4	45.45	60.30	3.53	8.23
5	46.45	61.63	3.94	8.73
6	43.40	63.63	3.11	10.57
7	47.27	64.70	4.00	10.57
8	49.77	65.03	5.71	11.20
9	48.43	65.00	5.17	11.00
10	56.22	65.23	6.61	11.20

Table 3 gives the mean *P* score on each trial. It is clear that the L12 items are more distinctive and easier to order. The L12 group starts off at a higher origin, reaches a higher asymptote, and improves more smoothly.

Representative serial position curves (Trials 1, 5, and 9) are shown in Fig. 1. The ordinate of Fig. 1 is the percentage of *S*s who placed a particular item in its correct position—irrespective of the placement of the other items. L12 begins with a serial position effect on Trial 1 and improves to nearly perfect ordering (Trial 9) where the serial position effect just about disappears. A serial position effect for L4 does not really emerge until quite late in learning. In fact, the later trials of L4 resemble most the earlier trials of L12. This can be seen from Table 4, which contains correlations of "percent correct" scores between L4 and L12, position by position, for representative trials.

The conclusion comes up again that ordering is retarded with L4. The high similarity of the items of

---

[3] Because *P* has this simple, direct interpretation, it seems better suited here than Kendall's tau. *P* is, however, directly related to tau. For an ordering of 12 items, $\text{tau} = \frac{P-33}{33}$.

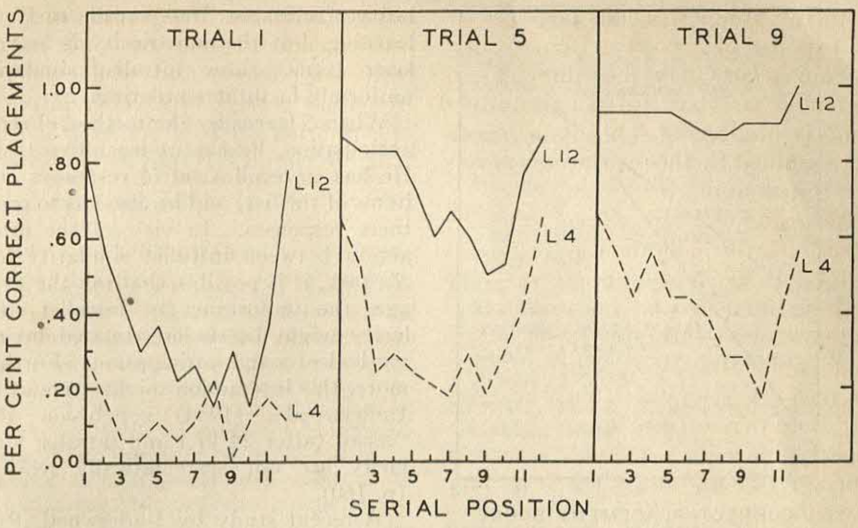


FIG. 1. Serial position curves on Trials 1, 5, and 9 for L4 and L12.

this list makes it extremely difficult for *S*s to discriminate sufficiently between the items to learn a unique ordering of them.

Another way to look at the serial position data is to ignore relative ordering and consider the number of items that *S*s placed in precisely the correct position on each trial. With this kind of measure—total number of correct placements—random arrangements of 12 items by a large population of *S*s would give essentially a Poisson distribution of scores (Feller, 1957, pp. 97–99) with a mean of 1. It can further be shown that if the precise location of k items is memorized and the remaining items are assorted randomly, the best estimate of the mean correct placement score would be $k + 1$.

The mean correct placement scores are listed in Table 3. Prior to Trial 8, the mean for L4 does not rise above 4.0. If *S* memorized the first, second, and last items and let the remaining items fall randomly, the expected mean would be 4.0. It is not until the mean score rises above this point (Trial 8) that a real serial position effect emerges. Thus, it seems that with the highly similar items of L4, the items are not ordered relative to one another, but are memorized individually.

If seven items are ordered 1, 3, 4, 5, 6, 7, 2 the ordering score is $P = 16$ (out of a maximum of 21) and the correct placement score is 1 (out of a maximum of 7). Thus, excellent relative ordering can be obscured by the correct placement score. In order to show how the two measures are related in empirical data, correct placement scores have been plotted against ordering scores in Fig. 2.

TABLE 4

CORRELATIONS, POSITION BY POSITION, BETWEEN "PERCENT CORRECT" SCORES ON L4 AND L12

L12	L4		
	Trial 1	Trial 5	Trial 9
Trial 1	0.39	0.84	0.80
Trial 5	—	0.73	0.90
Trial 9	—	—	0.62

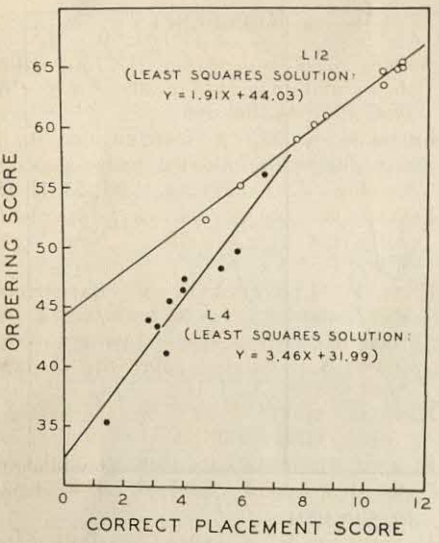


FIG. 2. Relationship between ordering scores and correct placement scores for L4 and L12.

For both lists the regression is linear, but the slopes differ considerably. If ordering improved as a function of correct placements only, about a three-point increase in ordering would occur to a one-point increase in correct placements. This is what happens with L4. When the correct placement score is 1, the ordering score is approximately 36. But the L12 ordering score starts considerably higher than one would predict from the 3:1 relationship. In fact, the L12 regression line suggests that had it been possible to measure performance early enough in learning, an ordering score of 46 would have corresponded to one correct placement. After *S* has reached a relatively good level of ordering, though, there is not much room for improvement, and the slope of the regression equation for L12 would have to be, and is, lower than that of L4.

These findings indicate the presence of an interaction between intralist similarity and *S*'s task. High intralist similarity facilitates free recall early in learning, but this superiority is lost on later trials. Low intralist similarity uniformly facilitates ordering.

When *S* learns by the method of serial anticipation, he has at least two tasks. He has to recall a set of responses (the items of the list) and he also has to order these responses. In view of the interaction between intralist similarity and *S*'s task, it is possible that, on the average, the importance of intralist similarity might be underestimated by the method of serial anticipation. Furthermore, this interaction might account for Underwood's (1954) conclusion that "recall (after 24 hr.) and intralist similarity are not systematically related" (p. 160).

A recent study by Underwood, Runquist, and Schulz (1959) indicates that paired-associate learning can be analyzed into two phases: (*a*) a response-learning phase, when responses become readily recallable, and (*b*) an associative phase, when each response is paired with a specific stimulus. One of their findings closely resembles a finding of this study: that high intralist similarity facilitates response learning. Since high intralist similarity hinders overall paired-associate learning (Feldman & Underwood, 1957), Underwood et al. (1959) inferred that the interference component of high intralist similarity occurs in the associative phase. In this respect the associative phase of paired-associate learning is like the ordering part of serial learning. This notion could be tested directly by teaching *S*s to match cards to appropriate stimuli without actually having them learn the responses per se. Thus, the method of the present study is directly applicable to the study of paired-associate learning.

The present results, furthermore, suggest the following way to maximize overall efficiency when *S* has to learn a number of lists by the serial anticipation method or the paired-associate method. The entire pool of items should first be grouped into lists of maximum intralist similarity for learning by the method of free recall. Then the items should be regrouped into lists of minimum intralist similarity for ordering or for associative learning. If there is no negative transfer from free recall to ordering or to associative learning, serial learning or paired-associate learning should then proceed very rapidly. This analysis needs to be tested experimentally.

SUMMARY

This study was designed to examine the joint influence of intralist similarity and the nature of *S*'s task on verbal learning. Two lists of 12 3-consonant trigrams were used. The trigrams of the high-similarity list contained only 4 different letters, while those of the low-similarity list contained 12 different letters. Two groups of *S*s recalled as many syllables as they could after each of 10 presentations, while two other groups learned the *order* of the syllables but not the syllables themselves.

The results indicate the presence of an interaction between intralist similarity and the nature of *S*'s task. High intralist similarity facilitates free recall early in learning, but this superiority is lost on later trials. Low intralist similarity, on the other hand, uniformly facilitates ordering. Since recall and intralist similarity have not been found to be systematically related when the method of serial anticipation was used, this interaction may be important in understanding the method of serial anticipation.

It was suggested that the method of this study might be applied to the study of paired-associate learning. The results also imply a way to maximize overall efficiency when *S* has to learn by the method of serial anticipation or by the method of paired-associates.

REFERENCES

ABORN, M., & RUBENSTEIN, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

FELLER, W. *An introduction to probability theory and its applications.* New York: Wiley, 1957.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N., & LEONARD, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1956, **51**, 1–11

KENDALL, M. G. *Rank correlation methods.* London: Griffin, 1955.

MILLER, G. A. Free recall of redundant strings of letters. *J. exp. Psychol.*, 1958, **56**, 485–491.

RUBENSTEIN, H., & ABORN, M. Immediate recall as a function of degree of organization and length of study period. *J. exp. Psychol.*, 1954, **48**, 146–152.

UNDERWOOD, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.

UNDERWOOD, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, **61**, 160–166.

UNDERWOOD, B. J., RUNQUIST, W. A., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

(Received May 4, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 58–61

FORWARD, BACKWARD, AND PSEUDOCONDITIONING OF THE GSR

R. A. CHAMPION AND J. E. JONES[1]

University of Sydney

For some time it has been customary to class forward and backward conditioning as fundamentally different processes and to regard the latter phenomenon as a form of pseudoconditioning (Underwood, 1949) or sensitization (Osgood, 1953) rather than true conditioning. While the evidence for this distinction is strong in the case of short-latency CRs (e.g., Spooner & Kellogg, 1947), there are theoretical reasons for believing that some true learning may occur when a backward conditioning procedure is used with a long-latency response (Mason, 1959). One reason for this expectation is that the presentation of the UCS prior to the CS by an appropriate time interval should allow greater contiguity of the CS and the UCR if the latter has a long latency. If it be supposed, however, that reinforcement is a necessary condition for learning as well as S-R contiguity then the backward conditioning procedure may mean greater delay of reinforcement and less propitious conditions for learning. Considering degree of S-R contiguity as the interval between the onset of the CS and the onset of the UCR, there may be greater contiguity in the backward procedure than in the forward procedure. On the other hand, assuming that proximity of reinforcement is the interval between the occurrence of the CR and the offset of the UCS, there may be greater proximity in the forward procedure than in the backward procedure. The optimum arrangement of stimuli for conditioning may be expected to vary with the latency of the response being conditioned, among other factors. In the present state of knowledge these optimum conditions cannot be accurately specified, it not being known, for example, that the reinforcement in classical aversive conditioning necessarily occurs with the offset of the noxious UCS, nor that any reinforcement at all is required in classical conditioning. As a first approximation the type of theory referred to above leads to the possibility of obtaining some degree of true backward conditioning of the GSR and the following data allow a test of this possibility. When the experiment was first designed it was assumed that a backward conditioning schedule would provide a valid control for pseudoconditioning, but this procedure produced evidence for true conditioning in the form of a clear and continued rise in performance and a group was therefore also tested with no short-term contiguity of CS and UCR, either forward or backward.

METHOD

Subjects.—The *S*s were 32 male and 28 female volunteers from undergraduate courses in psychology at the University of Sydney.

Apparatus.[2]—The *S* was seated in a sound-reduced room separate from that containing *E* and the main apparatus. The CS was a 2000-cps, 60-db. tone of 20 msec. duration presented through a speaker immediately behind *S*'s back. The ambient noise level in *S*'s room was approximately 50 db., being contributed to by the sound of an electric circulation fan. The UCS was a spike-wave ac shock of 90 cps, 3.20 milliamp., and 600 msec. duration. Shock strength was measured by means of an oscilloscope with a 15,000-ohm resistor in place of *S*. The shock was delivered to *S* through copper electrodes in the form of spring clamps 15 mm. wide which almost encircled the third and fifth fingers of the right hand near the fingertips. The GSR was picked up through thimble-type electrodes of German silver 30 mm. in length on the second and fourth fingers of the right hand. The shock and GSR electrodes were used in conjunction with electrode jelly. The GSR was measured and amplified with an electronic dc amplifier of the bridge type supplying constant current of 100 microamp. to the electrodes and followed by an ink-writing recorder. The time intervals listed above were determined by electronic interval timers, but when shock was used in isolation as a stimulus its duration was determined manually by *E* to avoid any possibility of giving cues to *S* and approximated 600 msec.

---

[1] Formerly Joan E. Mason and now at the Defence Research Medical Laboratory, Toronto, Canada.

[2] Part of the apparatus used in this study was provided by Commonwealth of Australia Research Grant No. 2002.

Procedure.—Following the receipt of instructions each *S* was given five trials with the UCS alone; the first and second shocks were 1.60 and 2.27 milliamp., respectively, and the remainder were 3.20 milliamp. The intertrial interval depended on *S*'s skin resistance reaching a steady level and was at least 60 sec.; the interval rarely exceeded 75 sec. and averaged about 65 sec. No ready signals were given at any stage of the experiment.

After the five preliminary UCS trials, Group TS ($N = 20$) was put on a forward conditioning schedule with 500 msec. between CS onset and UCS onset. Four CS-alone trials (T) were interspersed with the training trials (TS) as tests of conditioning and three more UCS-alone trials (S) were also included. The order of trials was TS, T, S, TS, TS, T, TS, S, TS, S, T, TS, TS, T. Group ST ($N = 20$) was put on a backward conditioning schedule with 750 msec. between UCS onset and CS onset; the order of trials was the same as for Group TS, but UCS-CS trials (ST) replaced the CS-UCS trials (TS).

After the data had been obtained from these two groups a third group of 20 *S*s was tested with a similar series of stimuli but with discrete presentations of tone and shock (Group T & S). The *S*s in this group received as many tones and shocks as the members of Groups TS and ST, but the stimuli were given in a different random order for each *S* and the minimum intertrial interval of 60 sec. was retained. The CS-alone trials (T) used as test trials in Group T & S were those which were preceded, on the average, by 6, 9, 13, and 15 shocks, as in the experimental groups. In addition, the whole series was begun with a CS-alone trial.

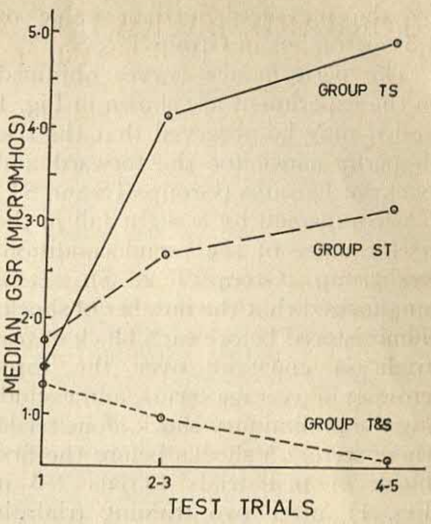


FIG. 1. Performance curves under forward conditioning (Group TS), backward conditioning (Group ST), and pseudoconditioning (Group T & S).

RESULTS

The procedure adopted with Groups TS and ST did not allow any measure to be taken of the GSR to the tone before any UCS-alone trials or CS-UCS trials. A number of experiments conducted in the University of Sydney laboratory have shown that the size of a GSR to a tone varies with the duration and intensity of the stimulus but that the median GSR before any training is approximately 1.5 micromhos for a 20-msec., 60-db. tone; a value of this order was therefore derived for each *S* in the two experimental groups by taking 15% of the mean score over the last three preliminary shock trials. The validity of this procedure is confirmed by the measured median value of 1.3 micromhos in Group T & S.

The performance curves obtained in the experiment are shown in Fig. 1 and it may be observed that the rise in performance for the forward and backward groups (Groups TS and ST) is accompanied by a slight fall in the performance of the pseudoconditioning group (Group T & S). It is emphasized that the number of shocks administered before each block of test trials is constant over the three groups; in average terms, and excluding the preliminary shock-alone trials, there were 2.5 shocks before the first block of test trials (Trials 2–3 in Fig. 1), after two training trials in Groups TS and ST, and nine shocks before the second block of test trials (Trials 4–5 in Fig. 1), after six training trials in Groups TS and ST.

Because the scores were not normally distributed, nonparametric tests were used. An *H* test (Edwards, 1954) was applied to the total GSR for the first and second blocks of test trials combined and it showed that the three groups differed significantly ($H = 27.79$, $P < .01$). This was followed by separate *U* tests (McNemar, 1955) comparing Groups TS and ST ($z = 1.39$, $P = .16$) and Groups ST and T & S ($z = 3.96$, $P < .001$). The trend in the three groups combined was examined over test trials by means of Mood's test (McNemar, 1955) and it proved significant ($\chi^2 = 7.30$, $P < .05$). Separate Mood tests then showed significant upward trends in Groups TS and ST ($\chi^2 = 10.90$ and 12.10, respectively, with $P < .01$) but no trend in Group T & S ($\chi^2 = 5.20$, $P > .05$). One other point of interest is that the difference between Groups TS and ST was significant on Test Trial 2 according to a *U* test ($z = 2.06$, $P = .04$). This is the stage at which the greatest rise in performance typically occurs in GSR conditioning (Champion, 1961; Moeller, 1954), as with many other responses.

DISCUSSION

The results point to a difference between backward conditioning and pseudoconditioning of the GSR as defined by the procedures used with Groups ST and T & S, respectively. The performance of Group ST showed a rise whereas that of Group T & S showed a nonsignificant fall. The initial improvement of the Group ST may not have been as great as that of the Group TS, but the two groups did not differ significantly over training as a whole. The general conclusion to be drawn is that the pseudoconditioning of the GSR is a basically different process from that occurring in forward and backward conditioning. A limitation to be placed on this conclusion at present is that the forward or backward interval between CS onset and UCS onset should not exceed 750 msec. In a pseudoconditioning control group Moeller (1954) used a backward conditioning procedure with the UCS-CS interval varying between 3000 and 5000 msec. He obtained no rise in performance over 23 trials on Day 2 of the experiment but did not report the effect of the 15 trials which all *S*s received on Day 1.

The present data lend weight to the possibility that UCS-CS trials may produce some degree of true conditioning in the case of a long-latency response such as the GSR. This outcome is in direct opposition to the prediction of Hilgard (1951) that a longer forward interval would be more appropriate with a long-latency response, which has been disproved (Moeller, 1954; White & Schlosberg, 1952). Further consideration of the initial argument on the definition of contiguity shows that a backward gradient of reinforcement must be assumed to operate in the conditioning of the GSR no matter what aspect of the UCS is taken as reinforcement unless the offset of the UCS is delayed an unusually long time. This is generally true in backward conditioning and applies in forward conditioning as long as the latency of the CR exceeds the interval between CS onset and UCS offset. If the concept of "backward reinforcement" is unacceptable then it presumably must be postulated that the conditioning of the GSR occurs without reinforcement. A possible reason for the failure to obtain true conditioning of a short-latency response with UCS-CS trials may be found if the appropriate latencies of the responses are considered. If the latency of the response is less than 500 msec., as with hand withdrawal or the eyeblink, then forward gradients of contiguity and reinforcement may be assumed in forward conditioning but both gradients must be taken as operating backwards in backward conditioning, again provided that UCS offset is not unduly delayed.

SUMMARY

A comparison was made of forward, backward, and pseudoconditioning of the GSR in the expectation that a backward procedure would produce true conditioning in such a long-latency response. The expectation was confirmed, for whereas the performance of the first two groups increased with paired CS-UCS or UCS-CS trials, that of the pseudoconditioning group decreased to a nonsignificant degree with comparable numbers of unpaired CS and UCS trials. Reasons were advanced for the failure on the part of other workers to obtain conditioning of a short-latency response with UCS-CS trials.

REFERENCES

CHAMPION, R. A. Supplementary report: Interpolated UCS trials in GSR conditioning. *J. exp. Psychol.*, 1961, in press.

EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

McNEMAR, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

MASON, J. E. The joint action of contiguity and reinforcement in classical conditioning. Unpublished doctoral dissertation, University of Sydney, 1959.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

SPOONER, A., & KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–334.

UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

(Received May 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 62–69

ACQUISITION OF PERCEPTUAL RESPONSES AS A FUNCTION OF LOADING, LOCATION, AND REPETITION [1]

HARRY W. KARN AND LEE W. GREGG

Carnegie Institute of Technology

The purpose of the present investigation was to determine the ways in which the visual perception of elements of a stimulus complex is influenced by the amount of experience with the complex and by specific features of the complex designated as location and loading.

The critical aspect of the experiment was the simultaneous presentation of three circles under spatial and temporal conditions which made the discrimination of a distinguishing binary cue (the presence or absence of a dot within each circle) difficult or impossible for all circles of the set. If the duration of the stimulus was made long enough, *S* could, by scanning the three circles, report accurately upon the nature of the individual circles. Given such an opportunity, *S* could learn that for one of the circles the value of the binary cue varied, while the values for the other circles remained constant over a series of presentations. These features of the stimulus field, if learned, could make it possible to report accurately at stimulus durations that precluded the possibility of scanning the entire stimulus field.

It was supposed that through such an experimental plan, evidence would be obtained to determine the ways perceptual responses are acquired by providing precise control over the nature of learnable cues and conditions of detection.

METHOD

Subjects.—The *S*s were 168 male college students, six in each of 24 experimental groups and 12 in each of two control groups.

Apparatus and procedure.—Three circles located at the points of an imaginary equilateral triangle were simultaneously presented to *S* on a screen 4 × 5 ft. One side of the triangle was 30 in. in length, subtending at *S*'s viewing distance approximately 15° of visual angle. The outside diameter of each circle was 1 in. corresponding to a visual angle of 2′ of arc. The discrimination required of *S* was the presence or absence, within each circle, of a dot subtending approximately 1′ of arc.

For the experimental *S*s all possible combinations of loading and location were used. Location refers to the position of the circle at the apex of the triangle (Location A), at one end of the base to *S*'s right (Location B), and at the other end of the base to *S*'s left (Location C). Loading refers to complete presence (positive loading), complete absence (negative loading), or random occurrence (random loading) of the dot within the circle at each location. The nature of the six configurations (Cs) presented to experimental *S*s are as follows: $C_1 = A+$, $B-$, Cr; $C_2 = A+$, Br, $C-$; $C_3 = A-$, Br, $C+$; $C_4 = A-$, $B+$, Cr; $C_5 =$ Ar, $B+$, $C-$; $C_6 =$ Ar, $B-$, $C+$.

Control *S*s observed circles in which the presence or absence of a dot was randomly determined for each of the three circles.

For each of the six Cs, four groups of experimental *S*s received 30, 20, 10, or 0 prepresentations, respectively, of the C at 1-sec. exposure followed by 60 presentations at .1-sec. exposure.

All *S*s were seated in an armchair to which was attached a panel with three toggle switches. Each switch had a neutral center position and could be thrown forward to indicate the presence of a dot or back toward *S* to indicate the absence of a dot. Small ink drawings of circles with or without the dot were placed in the appropriate directions for each switch. The switches were mounted in a triangular array to provide maximum S-R compatibility.

---

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-770.

The experimental room was dimly illuminated at all times by a single lamp located above *S*. The screen brightness was approximately 5 ml.

The instructions, which were read by *E*, informed *S* that three circles at the points of an imaginary equilateral triangle would be flashed on the screen before him and that sometimes one or more of the circles would contain a dot. A report on all circles was required with instructions to guess if unsure. No time limit for the response was imposed. After recording *S*'s responses by noting the order in which indicator lights came on, *E* asked *S* to return the switches to the neutral positions and to get ready for the next presentation.

RESULTS

The number of errors for each target for the prepresentations at 1 sec. was obtained for each *S*. Group mean errors at each location for the 36 *S*s of the experimental groups receiving 10, 20, or 30 prepresentations and for the 12 control *S*s who received 30 prepresentations are given in Table 1. The same error data for the experimental *S*s broken down according to the loading characteristics of the circles are also presented.

No errors were observed beyond the tenth presentation for any of the groups receiving preliminary experience at 1 sec. The mean total errors shown in Table 1 could therefore be considered as data from the first block of 10 prepresentations with 30 opportunities for error. It is clear from inspection of the table that regardless of location, loading, or the total number of prepresentations that *S*s finally received, no systematic variation in errors exists. The magnitude of the errors was uniformly low, about 10%, in every case.

TABLE 1

GROUP MEAN ERRORS DURING PRELIMINARY PRESENTATIONS AT 1 SEC.

Group	n	Location			Loading			Mean Total Errors
		A	B	C	+	−	r	
Exp. 10	36	1.3	1.1	1.1	1.2	1.2	1.1	3.5
Exp. 20	36	1.0	1.0	1.0	1.0	1.0	1.0	3.0
Exp. 30	36	1.1	1.2	.9	1.2	1.0	1.0	3.2
Cont. 30	12	1.2	1.1	1.0	random			3.3

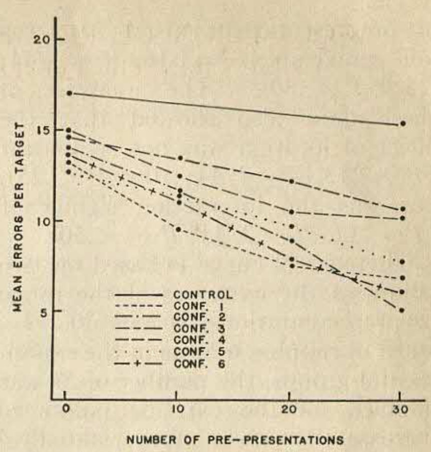


FIG. 1. The effect of preliminary experience on perceptual performance for six "loaded" configurations by experimental *S*s and an "unloaded" configuration by control *S*s.

For the 60 presentations at .1 sec. the number of errors for each target was obtained for each *S*. Group mean errors per target are shown in Fig. 1. The individual curves are distinguished by the particular combinations of loading at the three locations. These combinations were referred to earlier as Configuration (C). The solid curve of Fig. 1 connects the means of the two control groups.

It is apparent that preliminary practice had little effect on the performance of control *S*s. The mean for the Control 0 group is 16.92 and that for the Control 30 group is 15.11, a difference of 1.81 errors. An analysis of variance for the control group data indicated that the reduction in errors as a function of number of prepresentations at .1 sec. was not significant ($F = 1.16$, $df = 1/44$; $.25 < P < .50$). The analysis of these data also showed that the effect of location was not significant ($F = 2.13, df = 2/44; .10 < P < .25$), nor was the interaction significant ($F = .17, df = 2/44; P = <.50$).

The control curve is based on two points at the extremes of the range of prepresentations, 0 and 30. Instead of running 6 *S*s as in the experimental groups, the number of *S*s was doubled for the extreme points to increase the power of the statistical test. Based on the *S*s within error variance, the number of *S*s in each group required to detect a difference between the Cont. 0 and Cont. 30 means of 1.81 is 145.

For the *S*s of the experimental groups, there is a decrease in errors as a function of the variable, number of prepresentations. Separate analyses of variance were performed for the data of the experimental groups that had received the same C. Except for C_3, Number of prepresentations was a significant source of variation. The *F*s for this source in the analyses were 4.78, 8.27, 1.23, 5.67, 3.58, and 4.73, with 3 and 20 *df* for C_1 through C_6, respectively. The error variances for these analyses were homogeneous. The best estimate of the population variance for these treatments (36.52) was significantly smaller than that found in the control group analysis (58.80).

Loading and Location are necessarily confounded within any given configuration. The main effects of these variables can be demonstrated, however, by summing over Cs. The nature of these effects can be seen in the curves of Fig. 2. The left hand panel of this figure presents group mean errors per target as a function

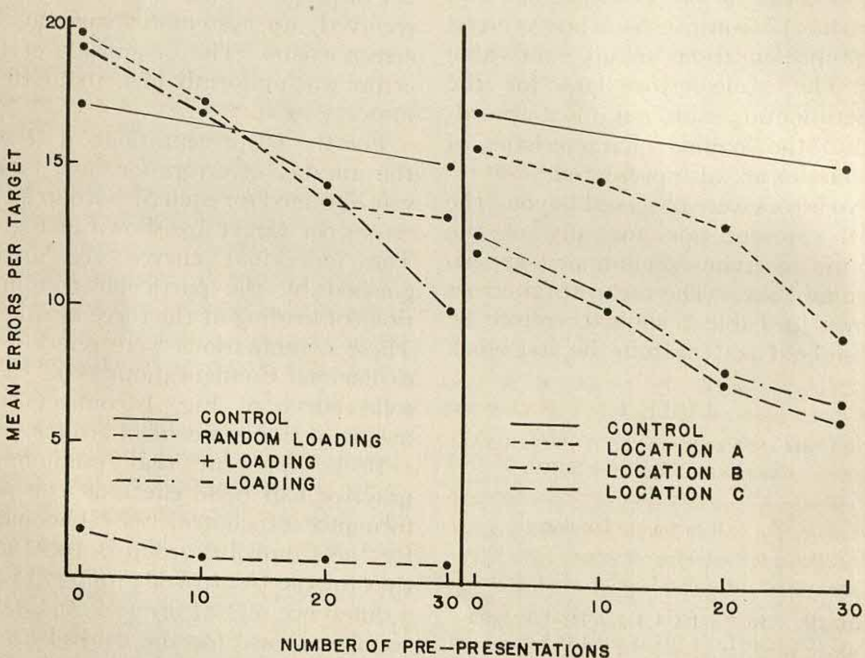


FIG. 2. Changes in perceptual performance as a function of preliminary experience with loading as the parameter (left panel) and with location as the parameter (right panel).

of the number of prepresentations at 1 sec. with Loading (regardless of Location) as the parameter. In the right hand panel of Fig. 2 the same data are used to compute group means with Location (regardless of Loading) as the parameter for the curves. In this figure the solid curves are based on the same means as the control curves of Fig. 1.

TABLE 2

Analyses of Variance of Errors per Target

Source	*df*	Configurations 1, 3, 5 Square 1		Configurations 2, 4, 6 Square 2	
		MS	*F*	*MS*	*F*
Loading (L)	2	5480	156.57*	4669	155.63*
Target location (T)	2	467	13.34*	478	15.33*
Configurations (C)	2	221	6.31*	123	4.10*
Square uniqueness	2	297	8.49*	699	23.30*
Number (N)	3	329	9.40*	547	18.23*
N × L	6	88	2.51*	106	3.53*
N × T	6	48	1.37	21	.70
N × C	6	17	.49	21	.70
N × U	6	25	.71	13	.43
Error_1 (*S*s within)	60	40	—	33	—
Error_2 (N × *S*s within)	120	32	—	28	—
Pooled error sources	180	35	—	30	—
Total	215				

* $P < .05$.

In order to assess the significance of the main effects of loading for the experimental groups, the data were treated as follows. Two latin squares were constructed so that the rows of the squares exhausted the set of six Cs. The first square contained the data for those experimental groups that received C_1, C_3, and C_5; the second square was for C_2, C_4, and C_6. For each square then, the columns represented Target Location; the rows represented Configurations as previously defined; and the letters represented Loading. It should be emphasized that although these squares were constructed, rather than randomly selected, together they provide the same information as the complete set of 12 3 × 3 latin squares. This is true because independent groups of *S*s were randomly assigned to the rows. Hence, row permutations of any kind cannot influence or bias the variance estimate and all possible column permutations are contained in the six Cs. The variable Number, refers to the prepresentations at 1 sec. and provides an additional source of variation orthogonal to the latin square sources. The summaries for the separate analyses are presented in Table 2. This form for presenting the outcome of the statistical treatment was selected in preference to a single combined analysis because it provides a comparison of what could be considered two independent sets of observations, each based on the performance of 72 *S*s. It should be noted that all basic sources of error are of the same magnitude. For each analysis the insignificant Error_1 (*S*s within groups) was pooled with the residual error, Error_2, to yield a single error term appropriate for all subsequent *F* tests.

The same inferences would follow from consideration of either of the analyses. The main effects of Loading, Target Location, Configurations, and Number of Prepresentations were significant beyond the .05 level in both.

Square uniqueness and the N × L interaction were also significant. Figure 2 is based on the combined data for all six Cs but nonetheless reflects accurately the origin of the N × L interaction. The curve for positive loading (dot always present) as a function of number of prepresentations is initially low and remains so. The group means are less than two errors per target for all levels of preliminary experience. The curves for the conditions of random and negative loading show a sharp and reasonably linear decrease. These

TABLE 3

Mean Errors per Target for Each Configuration, Loading, and Location

C	(+)	(−)	(r)	Mean
1	.7 (A)	11.4 (B)	15.7 (C)	9.3
3	.3 (C)	22.4 (A)	15.2 (B)	12.3
5	1.1 (B)	12.1 (C)	17.9 (A)	10.3
Mean	.7	15.3	16.3	10.6
2	.7 (A)	11.1 (C)	18.0 (B)	9.9
4	1.8 (B)	22.7 (A)	12.3 (C)	13.3
6	1.5 (C)	11.4 (B)	16.5 (A)	9.8
Mean	1.3	15.0	15.6	10.7

Note.—Each entry in Table 3 is the mean of 24 *S*s, six from the four experimental groups within each C. Using the pooled error estimates from both of the analyses of Table 2 gives a combined value of 32.5 as the σ_m per experimental unit. The $\sigma_{\text{diff.}}$ in the main body of the table is given by the expression $\sqrt{2ms/r}$. The computed value is 1.65. A difference as large or larger than 3.41 is significant at the .05 level.

marked differences in slope give rise to the significant interaction term.

The interpretation of the source called Square Uniqueness requires special comment. In the usual analysis of an unreplicated 3 × 3 latin square, the 8 *df* would be partitioned into 2 *df* each for row, column, and letter effects with 2 *df* remaining that may be used as an error term if there are no intrinsic interactions. Replicating a single square, as is the case for each of the squares in the present study, makes it possible to determine whether or not the particular square selected contains a unique pattern of confounded interaction effects if such interaction exists. In squares as small as the 3 × 3 one cannot suppose that confounded interaction effects would tend to cancel as they do in the larger squares. The significant Square Uniqueness effects in the present squares, therefore, means that there is some form of interaction among the variables Loading, Target Locations, and Configurations.

Table 3, which presents the means obtained by summing over *S*s of all four experimental groups receiving a particular C, makes it possible to account for the significant effect that Square Uniqueness indicates is present. In general, positive loading yields the fewest errors followed by the condition of negative loading with the random loading condition producing the greatest number of errors. This pattern obtains for C_1, C_2, C_5, and C_6. C_3 and C_4 depart radically from this pattern, reversing the error magnitudes associated with negative and random loadings. Moreover, the means for the negative loadings are the largest in the entire table. C_3 and C_4 have one thing in common—the negatively loaded circles are at Location A, the apex of the triangle.

The preceding analyses indicate that without specific instructions to do so *S*s learn to utilize the cues inherent in the loaded fields to improve the accuracy of their perceptual reports. This learning is a function of the amount of prior experience with the target field.

Does proficiency of performance change during the presentations at the .1-sec. exposure even without prior experience? Are the rate and extent of such changes dependent upon amount of prior experience?

To answer these questions blocks of 20 consecutive presentations within the 60 presentations at .1 sec. provided three error measures for each *S*. These measures represent total errors for all targets and can take on values from 0 to 60 errors. Means based on 36 *S*s in the experimental groups and 12 *S*s in the control groups are plotted in Fig. 3 as a function of the ordered blocks of presentations. Number of prepresentations at 1 sec. is the parameter.

Summary tables for the separate

analyses of variance for the control and experimental groups are presented in Table 4. The main effects of Number of prepresentations and Configurations are significant in the case of the experimental groups. Number is not significant as a source of variation in the control group analysis. These main effects have their counterparts in the latin square analyses and control group analysis mentioned earlier.

The main effect of Blocks is of primary concern in the present analyses. For the experimental groups, the Block effect and the B × N interaction are significant. Error reduction occurs, and does so at a differential rate depending upon the number of prepresentations, within the 60 presentations at .1 sec. For the control groups, neither the main effect of Blocks nor the B × N interaction was significant at the .05 level.

When the number of prepresentations at 1 sec. is 0, the sharpest drop in errors occurs over the blocks of presentations at .1 sec. With increasing numbers of prepresentations, the rate of error reduction within the 60 presentations at .1 sec. decreases as indicated by the slopes of the curves for the experimental groups in Fig. 3.

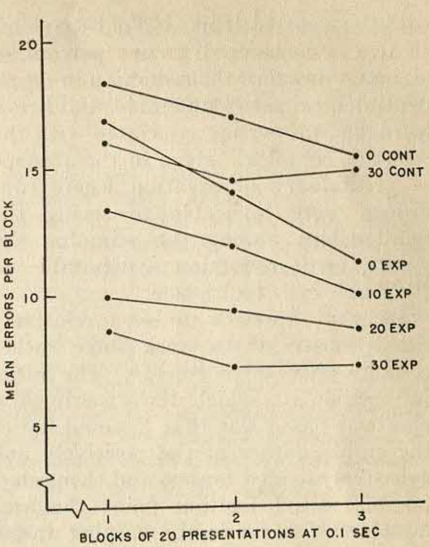


FIG 3. Changes in perceptual performance within 60 trials at .1 sec. for control and experimental *S*s receiving differing amounts of preliminary experience.

DISCUSSION

Under the conditions of the present experiment, a reduction in errors in reporting the state of three simultaneously presented targets occurred. The magnitude of the error reduction was a function of amount of prior experience

TABLE 4

ANALYSIS OF VARIANCE OF ERRORS PER BLOCK

Source	Experimental Groups			Control Groups		
	df	*MS*	*F*	*df*	*MS*	*F*
Number (N)	3	863.42	23.68*	1	58.68	.99
Configurations (C)	5	136.56	3.75*			
N × C	15	18.40	.50			
Error_1 (*S*s within)	120	36.46	4.52*	22	58.80	6.30*
Blocks (B)	2	291.88	36.21*	2	28.35	3.04
B × N	6	47.32	5.87*	2	7.68	.8
B × C	10	6.34	.79			
B × N × C	30	3.63	.45			
Error_2	240	8.06	—	44	9.33	—
Total	431			71		

* $P < .05$.

with the targets. This kind of behavioral change is considered to be "perceptual learning" because the reduction in errors depended primarily on stable, and hence learnable, properties associated with the stimulus complex. Even in the absence of preliminary observation where conditions were favorable to seeing the relationships among the stimulus elements, error reduction attributable to the stable cues took place.

In what appears to be a relatively simple observational task some rather complex phenomena emerge. The initial supposition on which the experimental plan was based was that *S* would notice the stable nature of the positively and negatively loaded targets and then select for his initial fixation point the area in which the randomly varying target would appear. In this way nearly perfect performance could be expected. Moreover, having seen this solution, *S*'s response patterns should show the sudden, insightful shift typically associated with problem solving tasks. Such was not the case. Instead, the improvement, in general, was slow and incremental for the negatively loaded and randomly varying targets. The only marked discontinuity was the extremely rapid reduction in errors for the positively loaded target. This superior performance with respect to the positively loaded target coupled with the relatively high number of errors for the negatively loaded target, especially when the latter occupied a nonpreferred position in the stimulus field, poses a paradox since both elements provide the same degree of stability as a cue for learning.

It is evident that a complete explanation of the observed relationships will depend upon processes other than those implied in the original formulation of the task. Certainly, *S*s will and do acquire information about the relationships that exist within the stimulus complex. This information provides the background against which the perceptual responses to the complex are made. However, even in an arrangement as limited as the three binary-valued elements used in this study, several additional cognitive and response processes must play a role.

First, consider the habitual ways of scanning the stimulus field (Carmichael & Dearborn, 1947). When more than one element is simultaneously presented, attention will be divided, shifting element to element if time allows. Depending on the internal relationships within the stimulus complex, a variety of possible perceived relationships could be educed as a function of the relative positions of the elements. Or what of the effect of set? Perhaps a partial description of the perceptual learning phenomena reported here rests in the establishment of perceptual sets akin to the cognitive sets in certain problem solving experiments (Gardner & Runquist, 1958; Kendler, Greenberg, & Richman, 1952). Of a similar nature is the finding that response sets can be established which operate in the absence of any physical stimulus (Goldiamond & Hawkins, 1958). Not unlike the response sets are the factors related to *S*'s criteria for reporting and these, as has been suggested, may interact with the more basic variables such as intensity which determine the probability of seeing (Tanner & Swets, 1954).

At present, full-scale investigations of various instructional sets and of the "probability of seeing" effects on the kind of perceptual learning reported here are being conducted. Until such a time as perceptual theory can bring together in coherent sequence the complex processes that generate such behaviors, the present data should stand as a caution to investigators who elect to present tachistoscopically complex figures, words, and the like and from such presentations draw inferences concerning their perceptual consequences.

SUMMARY

The aim of this study was to determine the ways perceptual responses are acquired by providing precise control over the nature of learnable cues and conditions of detection.

Three circles located at the points of an imaginary equilateral triangle were simultaneously presented to *S*s on a screen by means of a tachistoscope with instructions

to report presence or absence of a dot within each circle. Loading refers to complete presence, complete absence, or random occurrence of the dot at a location.

For each of six possible combinations of loading and location, four groups of experimental *S*s received 60 presentations at .1-sec. exposure following either 0, 10, 20, or 30 preliminary presentations at 1 sec. Control *S*s observed circles in which the presence or absence of a dot in all circles was randomly determined. Since conditions were such that it was impossible to observe all circles at .1 sec., changes in performance may be attributed to experience at the 1 sec. exposure.

There was a significant reduction in errors as a function of the number of preliminary 1-sec. presentations under all conditions of loading and location. But both location and loading influenced the magnitude of the error, and there was a significant interaction between loading and location.

Tentative explanations of the obtained differential loading and location effects include habitual scanning procedures, perceptual sets, and subjective criteria used in reporting.

REFERENCES

CARMICHAEL, L., & DEARBORN, W. H. *Reading and visual fatigue.* Boston: Houghton Mifflin, 1947.

GARDNER, R. A., & RUNQUIST, W. N. Acquisition and extinction of problem-solving set. *J. exp. Psychol.*, 1958, **55**, 274–277.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

KENDLER, H. H., GREENBERG, A., & RICHMAN, H. The influence of massed and distributed practice on the development of mental set. *J. exp. Psychol.*, 1952, **43**, 21–25.

TANNER, W. P., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

(Received May 10, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 70–79

ASSOCIATIVE FACTORS IN VERBAL TRANSFER[1]

JARVIS BASTIAN[2]

University of Minnesota

In the two decades since the publication of Gibson's (1940) well-known analysis of verbal transfer in terms of generalization and differentiation, an impressive amount of research has been directed towards tracing out the effects of verbal generalization in a wide range of transfer situations. These efforts have produced a body of results of sufficient breadth and uniformity to permit the construction of fairly general statements of the experimental effects of verbal generalization, such as those provided by Osgood (1953) and Underwood (1949).

However, comparatively little has been done to determine the underlying processes responsible for these effects. It is generally conceded that verbal generalization, aside from that which is ascribable to the phonetic or orthographic properties of the experimental materials, is the result of processes quite different from those governing primary generalization. Nevertheless, the feeling persists that there must be parallels between the two phenomena, particularly in regard to the roles played by similarities in the verbal domain and the physically specifiable similarities associated with primary generalization.

It is perhaps for this reason that the notion of "meaning" is so often invoked in theoretical discussions of verbal generalization, because "similarities in meaning" apparently provide a basis for verbal generalization that is analogous to the similarities in physical attributes involved in primary generalization.

In spite of the extremely formidable logical and linguistic problems that are involved in the notion of "similarities of meaning," its use in this context has been strengthened by a considerable amount of experimental evidence clearly relating the degree to which verbal materials are judged to be similar in meaning to the degree of generalization obtaining between these materials. Haagen (1943), Morgan and Underwood (1950), and Underwood (1951) have all reported transfer results in paired associate training which indicate that generalization effects increase as a function of the degree to which words in comparable interlist positions are judged to be similar in meaning.

However, the results of these studies are possibly contaminated by another, uncontrolled, factor in the experiments. The experimental variable, degree of similarity in meaning, was manipulated in all of these experiments by selecting words from a set of 400 pairs of adjectives for which Haagen (1949) had collected ratings of degree of synonymity. But Haagen also obtained, from a separate group of judges, ratings of the degree of associative connection between these same word pairs. The correlation between the scale values derived from the two sets of ratings was .90, which was only slightly less than their individual reliability coefficients. Thus, in the experiments cited, the generalization results can be attributed as much to the associative factor as to the similarities in meaning, if, indeed, these appear as separate factors in these studies.

Although degree of semantic similarity and degree of associative connection are evidently highly correlated, it might be possible to manipulate them separately because of the considerably broader range of interverbal relations covered by associative tendencies. The exploration of this possibility is highly desirable, for McClelland and Heath (1943) and Russell and Storms (1955) have reported transfer experiments which suggest that associative factors may be fundamentally involved in the production of verbal generalization. These studies both demonstrated pronounced generalization effects across paired associate lists containing words in analogous interlist positions which were linked by associative tendencies, as inferred from word association norms, but which were not semantically related in any acceptable sense.

---

[1] This report is based on part of a PhD dissertation presented to the Graduate School of the University of Minnesota. The author is deeply indebted to W. A. Russell and J. J. Jenkins for their advice and encouragement.

[2] Now at Haskins Laboratories, New York, New York.

The present experiment was designed to assess the separate and combined generalization effects produced by the two different interverbal relations in order to provide some further information that might lead to a clearer conceptualization of the factors mediating verbal generalization.

METHOD

The strategy of this research required a situation that provided a sensitive index of verbal generalization, and one in which the factors of associative strength and similarity of meaning could be separated and their effects compared. For this purpose a transfer of training situation was selected that involved the successive learning of two lists of paired associates containing identical stimulus elements. The basis for the selection of this experimental situation is that in the successive learning of paired-associate lists, each containing the same set of stimulus items, but different response items, generalization between the responses results in positive transfer, whereas the lack of generalization results in negative transfer (cf. Osgood, 1953; Underwood, 1949). In addition, parallel experiments conducted by Haagen (1943) have indicated that this situation is more sensitive to generalization effects than those in which responses are held constant across lists and stimuli varied, or those in which both stimulus and response elements are varied.

Materials.—With these considerations in mind, lists of 12 paired associates were constructed with the same set of words appearing as stimulus elements in each. The response elements in List 1 were taken from the stimulus words used in the Kent-Rosanoff free association test.

Each response element in the other lists was related to a given response element in List 1 by one of the following three relationships: Type A elements were highly frequent word associates of the List 1 response words, but were judged to bear little or no similarity to them in meaning; Type S elements were very infrequent word associates of the List 1 response words, but were judged to be highly similar to them in meaning; Type C elements were control words which were neither associates of, nor judged to be similar in meaning to the List 1 response words.

The measure of frequency of association between the List 1 response words and the response words in the other lists was obtained from the frequency with which these words appeared as responses to the List 1 words in the Kent-Rosanoff norms provided by Russell and Jenkins (1954). The measure of similarity of meaning between the List 1 response words and those in the other lists was obtained from median judgments provided by a group of 32 judges not otherwise pariticipating in the experiment. The judges were asked to rate each pair of words on a four-point scale ranging from a minimum at "0" (defined for them as: "no similarity or even opposite in meaning") to a maximum at "3" (defined as: "identical or very similar in meaning").

From a much larger pool of word pairs for which these measurements were available, 36 words were selected which most closely matched the properties of each of the Type A, S, or C relationships to the List 1 response words. (It turned out to be exceptionally difficult to discover words which met these combined requirements. Evidently, if a pair of words in general use is judged to be similar in meaning, it is also quite likely that each will be evoked by the other in a word association test.) Lists 2, 3, and 4 were then constructed by pairing each of these words with the same stimulus word that had been paired with their counterparts in List 1, so that Lists 2, 3, and 4 each contained four response words representing each of the three types of interlist relationships. The composition of the lists and the type of relation each of the response words in Lists 2, 3, and 4 bore to the List 1 response words are presented in Table 1.

TABLE 1

Composition of Lists and Interlist Relations

Stimulus Words All Lists	List 1 Responses	List 2		List 3		List 4	
		Responses	Type	Responses	Type	Responses	Type
BREAD	HIGH	LOFTY	S	LEFT	C	LOW	A
COLD	STEM	TRUNK	S	FLOWER	A	PROVE	C
EAGLE	SICKNESS	HEALTH	A	BROOK	C	DISEASE	S
FRUIT	QUIET	LOUD	A	HAND	C	SILENCE	S
HAMMER	SOUR	SWEET	A	ACID	S	NARROW	C
JUSTICE	DREAM	IMAGINE	S	MARKET	C	SLEEP	A
MUTTON	LONG	SAND	C	SHORT	A	TALL	S
NEEDLE	KING	QUEEN	A	CHIEF	S	CLOSE	C
OCEAN	BOY	NORTH	C	LAD	S	GIRL	A
SLOW	LIGHT	MINE	C	DARK	A	GAY	S
TABLE	MAN	CONTAIN	C	HUMAN	S	WOMAN	A
THIRSTY	HARD	DIFFICULT	S	SOFT	A	LITTLE	C

All words were carefully screened to eliminate any other associative or semantic relations within or between the lists, and to eliminate orthographic similarities and words having low frequencies of usage.

In order that differences in the extent of generalization might be confidently inferred from differences in performance in the transfer task, it would first be necessary to establish that such differences in performance could not be attributed to intrinsic differences in the ease with which different types of pairs are learned, irrespective of the effects of the prior learning of the related pairs. For this reason it was desirable to have different groups of *S*s learn the lists in two different sequences (Sequences I and II) so that the subsequent task in one sequence would be the same as the prior task in the other sequence. It was also thought desirable to investigate the generalization effects produced by the various interlist relations in the recall of the original pairs. Therefore, the order of tasks presented in Sequence I was: learn List 1; learn List 2, 3, or 4; recall List 1. The comparable order in Sequence II was: learn List 2, 3, or 4; learn List 1; recall List 2, 3, or 4.

However, it should be appreciated that the associative relation is not necessarily sym-

TABLE 2

Percentage Association and Median Ratings of Similarity to List 1 Response Words for Words in Each Type of Interlist Relation

Type A	Association		Similarity	Type S	Association		Similarity	Type C	Association		Similarity
	Seq. I	Seq. II			Seq. I	Seq. II			Seq. I	Seq. II	
DARK	64.2	82.2	0.53	ACID	0.2	6.5	1.92	BROOK	0.0	0.0	0.50
FLOWER	39.9	1.0	0.84	CHIEF	0.1	0.0	2.38	CLOSE	0.0	0.0	0.52
GIRL	76.2	69.8	0.57	DIFFICULT	0.5	65.5	3.11	CONTAIN	0.0	0.0	0.53
HEALTH	37.3	24.8	0.53	DISEASE	0.1	19.4	2.44	HAND	0.0	0.0	0.50
LOUD	34.5	6.7	0.52	GAY	0.0	0.0	2.07	LEFT	0.0	0.0	0.50
LOW	67.0	64.0	0.53	HUMAN	0.3	36.0	1.65	LITTLE	0.0	0.0	0.50
QUEEN	74.5	71.0	0.99	IMAGINE	0.0	22.3	2.53	MARKET	0.0	0.0	0.53
SHORT	75.2	33.3	0.52	LAD	0.5	69.1	3.16	MINE	0.0	0.0	0.50
SLEEP	44.9	4.7	0.94	LOFTY	0.1	64.0	2.99	NARROW	0.0	0.0	0.52
SOFT	66.9	44.1	0.52	SILENCE	0.6	53.2	3.27	NORTH	0.0	0.0	0.50
SWEET	56.3	43.1	0.53	TALL	0.5	0.7	1.82	PROVE	0.0	0.0	0.53
WOMAN	76.1	61.1	0.57	TRUNK	0.7	0.0	2.25	SAND	0.0	0.0	0.50
Mean	59.4	42.8	0.63		0.4	28.1	2.47		0.0	0.0	0.51

metric. That is, the extent to which Word A evokes Word B as an association response is not necessarily the same as the extent to which Word B evokes Word A. Thus it was necessary to determine the frequency with which the response words in List 1 tended to be given as association responses to their counterparts in List 2, 3, or 4. This frequency was inferred from additional association norms collected by Russell and Jenkins (1956) and from norms collected by *E* from 139 *S*s not otherwise participating in the experiment.

Table 2 presents the interlist association and similarity properties in both sequences for words in each type of interlist relation. It can be seen that on the average the properties of the interlist relations in the converse order (Sequence II) were different from those obtaining in Sequence I in that the response words participating in the Type S relation were characterized by a substantial degree of interlist associative linkages, whereas the associative linkages between the words of Type A were somewhat weaker than those in Sequence II. Since of necessity, the similarity in meaning relation is symmetric, the words representing the Type S interlist relation in Sequence II combined to a considerable extent the properties of both Type A and Type S relations.

Procedure.—Seventy-two *S*s, students in an introductory psychology course, participated in the experiment, 36 learning the lists in each sequence. In order to obtain a measure of the effect of the different lists, *S*s were further divided into three equal groups within each sequence, according to the particular lists they were to learn during the different phases of the experiment. In Sequence I *S*s in Groups 1, 2, and 3 learned Lists 2, 3, and 4, respectively, as their second lists, and in Sequence II Lists 2, 3, and 4 were learned as their original lists by Groups 1, 2, and 3, respectively.

After a short period of pretraining on a practice list of nonsense syllable-digit paired associates, *S*s proceeded to learn two of the experimental lists by the anticipation method, with instructions to pronounce aloud both stimulus and response words on all occasions. The materials were presented on a Hull-type memory drum arranged so that the stimulus words were exposed for 2 sec. and then were joined by the response words for another 2 sec. Two seconds intervened between the presentation of each pair, and 8 sec. separated successive trials. The 12 pairs in each list were presented in five different orders.

The *S*s were instructed to correct all mistakes aloud and to continue practice on the lists until all response words had been correctly anticipated in a single trial. When this criterion had been reached on the original list, a shutter arrangement on the apparatus was shifted to expose the second list, and after a brief period for instructions lasting approximately 1 min., practice began immediately on the second list and continued until the same criterion was met. Then, immediately following an instructional period of the same duration in which *S*s were told to recall as many of the original responses as possible on the first trial, the shutter arrangement was shifted back to the original list, and the recall test began and continued until all responses had been correctly anticipated in one trial.

RESULTS

Original list learning.—The number of trials required to reach the criterion of correctly anticipating all responses in the original lists was analyzed for any overall differences in the ease with which the lists were learned, irrespective of the types of interlist relations in which their component responses would subsequently participate. The mean numbers of trials needed to attain this criterion for Groups 1, 2, and 3 in learning List 1 in Sequence I were 13.17, 11.83, and 12.67, respectively. The groups learning Lists 2, 3, and 4 in Sequence II required 14.38, 11.92, and 11.00, respectively. The differences among the group performances on these lists cannot be considered reliable ($F = .92$, $df = 2/66$), nor were the other sources of variation significant.

The total numbers of correct anticipations given by all *S*s in learning the original lists in each sequence were analyzed for differences among the different types of interlist relations. Since each list contained words from each category, measurements of *S*s' performance on them cannot be considered independent. Therefore, Collier's (1958) modification of the model for the analysis of observations correlated on one dimension presented by Kogan (1948) was applied.

The mean total correct anticipations of response words of Types A, S, and C, respectively, were 31.69, 31.89, and 32.39 in Sequence I, and 32.00, 31.91, and 31.08 in Sequence II, and 31.84, 31.90, and 31.73 for the two sequences combined. The differences among these means cannot be considered reliable, nor were any reliable differences found between groups or sequences (all Fs $< .80$).

However, significant Type×Group interaction effects did appear in the combined sequence analysis ($F = 15.11$, $df = 4/132$, $P < .01$) and in the analyses of each sequence separately. For Sequence I, in which the groups were all learning List 1, $F = 10.53$, $df = 4/66$, $P < .01$, and for Sequence II, $F = 7.08$, $df = 4/66$, $P < .01$. The appearance of this effect, even when the groups were learning the same list, suggests that it was due to the considerable differences in the ease with which individual paired associates were learned. In Sequence I, the difficult pairs in List 1 were variously classified as Types A, S, and C, according to the interlist relations they would form when the groups subsequently learned the other lists. These differences in learning the individual paired associates were somehow related to the stimulus components of the lists, since certain pairs sharing the same stimulus words were uniformly more difficult to learn in all the lists.

This same interaction also appeared in analyses of the number of trials required to reach the criterion of correctly anticipating all responses of a given type on a single trial ($F = 8.82$, $df = 4/66$, $P < .01$ in the combined sequence analysis). The mean trials required to reach this criterion for Types A, S, and C, respectively, were 8.34, 8.95, and 9.23 in Sequence I, and 8.92, 8.26, and 9.34 in Sequence II. Neither these differences, nor those due to other sources in the analysis, can be considered reliable (all Fs < 1.80). Thus any type differences in performance appearing when these lists were presented in the transfer test can be assumed to be the differential effect of previous learning.

Second list learning.—The possibility that the use of *S*s' performance in learning the original list as a covariance control might effect greater precision in the analysis of the results of the transfer and recall tests had to be rejected, as the regression coefficients between original list performance and the other measures were uniformly small and could not be regarded as reliably different from zero.

The number of trials required to reach the criterion of correctly anticipating in a single trial all responses in the second lists, regardless of type memberships, was analyzed for overall differences in the learning of the lists. For the groups learning List 1 as their second list in Sequence II, 9.83, 8.67, and 10.50 trials were required, respectively. The groups learning Lists 2, 3, and 4 in Sequence I needed 10.67, 9.00, and 9.00 trials, respectively. These group performances cannot be regarded as different ($F = .58$, $df = 2/66$), nor were the other sources of variance significant.

Differences in verbal generalization related to differences in interlist relations have typically been found to be greatest in the initial stages of second list learning and original list relearning (cf. Haagen, 1943; Morgan & Underwood, 1950; Osgood, 1946, 1948; Underwood, 1951). Therefore, performances on the first anticipation trial on the second lists were examined.

TABLE 3

ANALYSIS OF VARIANCE OF PERFORMANCES IN LEARNING SECOND LISTS

Source	df	Correct Anticipations First Trial		Correct Anticipations All Trials		Trials to Criterion	
		MS	*F*	*MS*	*F*	*MS*	*F*
Within *S*s	144						
Types	2	31.60	43.89**	753.50	48.30**	71.30	10.45**
Types × Seq.	2	1.85	2.57	173.40	10.45**	34.90	5.12**
Types × Groups	4	3.72	5.17**	452.55	29.01**	76.30	11.19**
Types × Groups × Seq.	4	.90	1.25	9.98	.64	1.10	.16
Error 1	132	.72		15.60		6.82	
Between *S*s	71						
Sequences	1	1.50	1.35	65.50	.16	10.20	.30
Groups	2	2.20	1.98	216.05	.52	50.10	1.46
Seq. × Groups	2	3.05	2.75	117.70	.28	30.30	.88
Error 2	66	1.11		418.22		34.31	

** $P < .01$.

The results of these analyses are presented in Table 3.

A Type × Group interaction effect appeared in this analysis as it had in analyses of performances in the original learning of these same lists. The different types of interlist relations were also an appreciable source of variation.

In order to compare the Type means in the two sequences, Snedecor's (1956) procedure for testing comparisons among all means was followed, using the variance estimate from the combined sequence analysis to establish confidence intervals about the means. This indicated that all differences within and between sequences were significant at the .05 level, except for the difference between the Type A means in the two sequences, and the difference between the Type A and S means in Sequence II, which could not be considered reliable. The Type means for each sequence are presented in Table 4.

In order to examine the effect of the interlist relations beyond the first trials, an analysis of the total correct anticipations produced throughout the course of learning the second lists was performed. In addition, the number of trials required to reach the criterion of correctly anticipating all words of a given type within a single trial was also analyzed. The results of these

TABLE 4

PERFORMANCES ON WORDS OF EACH TYPE IN LEARNING SECOND LISTS

Type	Correct Anticipations First Trial				Correct Anticipations All Trials				Trials to Criterion			
	Sequence I		Sequence II		Sequence I		Sequence II		Sequence I		Sequence II	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A	1.97	.99	1.78	.97	33.00	11.97	31.42	13.68	5.56	3.82	6.56	4.60
S	1.27	.94	1.56	1.23	30.39	11.08	30.78	12.71	6.64	4.34	6.12	4.09
C	.36	.63	.78	.26	23.72	12.14	28.22	12.23	8.81	3.59	7.23	4.16

TABLE 5

Analysis of Variance of Performances in Recalling and Relearning Original Lists

Source	*df*	Correct Anticipations Recall Trial		Correct Anticipations All Trials		Trials to Criterion	
		MS	*F*	*MS*	*F*	*MS*	*F*
Within *S*s	144						
Types	2	17.00	20.24**	59.60	4.53*	12.85	3.34*
Types × Seq.	2	.20	.24	4.20	.32	1.20	.31
Types × Groups	4	6.75	8.04**	94.40	7.18**	29.78	7.74**
Types × Groups × Seq.	4	.90	1.07	6.20	.47	2.72	.71
Error 1	132	.84		13.15		3.85	
Between *S*s	71						
Sequences	1	.00	.00	.00	.00	17.20	1.28
Groups	2	2.60	1.50	156.10	.48	10.90	.81
Seq. × Groups	2	.50	.29	32.05	.10	5.45	.40
Error 2	66	1.73		323.31		13.48	

* $P < .05$.
** $P < .01$.

analyses are also presented in Table 3.

In both analyses the Type effect continued to be significant. Again, the Type × Group effect appeared in each, and the Type × Sequence effect also attained significance in these analyses. The Type means within each sequence in both analyses are presented in Table 4.

Applying Snedecor's procedure to the means of the total correct anticipations produced on all trials revealed that all Type differences within each sequence were significant, except for that between Types A and S in Sequence II. All Type differences between sequences were significant except for those between Type S in Sequence I and Types A and S in Sequence II.

In the analysis of the trials required to reach criterion, all differences within Sequence I were found to be significant, but in Sequence II Types A and S were different from Type C but not from each other. Between sequences, all differences were reliable except for those between Type S in Sequence I and Type A in Sequence II and between the two Type S means.

Recall and relearning of original lists.—The relearning of the original lists was a comparatively easy task for all *S*s. The numbers of trials needed in Sequence I to reattain criterion on List 1 were 4.75, 5.00, and 6.08 for Groups 1, 2, and 3, respectively, while in Sequence II the groups relearning Lists 2, 3, and 4 required 5.92, 6.17, and 7.50 trials, respectively. The differences between these groups cannot be considered significant ($F = .89$, $df = 2/66$), nor were the other sources of variance significant.

The analyses of the number of correct anticipations for words of each type on the recall trial are presented in Table 5. Both Type and Type×Group interaction effects were found. Within both sequences Types A and S differ significantly from Type C, but not from each other. None of the Type A and S differences between sequences were reliable, but each was significantly different from either Type C means. The type means from both sequences are presented in Table 6.

Performance on relearning the original lists was measured by the total correct anticipations given on all trials and by the number of trials required to reattain criterion. Analyses of these performances are presented in Table 5. Both Type and Type × Group effects appeared in each analysis. The Type means in each Sequence obtained from analysis of both measures of relearning are presented in Table 6.

In the analysis of the number of correct anticipations produced throughout relearning, it was found that, within both sequences, Types A and S were both reliably different from Type C, but not from each other. As in the recall test, none of the Types A and S differences between sequences were reliable, but each was itself significantly different from either Type C means.

The analysis of the number of trials required to reach the relearning criterion in Sequence I indicated that the Type A and S means were each reliably different from the Type C mean, but not from each other. None of the differences among the Sequence II means were significant. Type A in Sequence I differed from all Sequence II means and Type S in Sequence I differed from Types A and C in Sequence II, but none of the other between-sequence differences were reliable.

DISCUSSION

All analyses of second-list learning and first-list recall and relearning clearly indicate that associative interlist relations are sufficient to produce verbal generalization. Indeed, the results of the most sensitive test of interlist transfer indicate that words participating in associative relations, irrespective of any attendant semantic affinities, were more easily anticipated than words linked only by semantic relations. This suggests that associative relations may even be a necessary factor in interverbal generalization, and that words having semantic relations typically yield generalization effects only because they are usually associated or that they are particularly susceptible to being temporarily associated.

If this were so, the manner in which associative factors operate might be as follows. At the start of learning the pairs in the transfer tests, the responses most likely to occur upon the initial presentations of the stimulus words are those which have been paired to these same words in the initial lists. But because of the establishment of associative connections prior to the experiment, there exists a substantial tendency for these responses, intruding from the preceding lists, to evoke the correct responses in the transfer tests. Hence, these presumably implicit interlist intrusions facilitate the learning of the second list, as their occurrences act as mnemonic aids to learning the transfer list.

TABLE 6

PERFORMANCES ON WORDS OF EACH TYPE IN RECALLING AND RELEARNING ORIGINAL LISTS

Type	Correct Anticipations Recall Trial				Correct Anticipations All Trials				Trials to Criterion			
	Sequence I		Sequence II		Sequence I		Sequence II		Sequence I		Sequence II	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A	2.94	1.00	2.81	1.02	16.28	9.27	15.86	9.81	2.86	2.00	3.78	3.06
S	2.69	1.02	2.69	1.10	16.36	9.91	16.25	9.93	3.03	1.74	3.64	3.36
C	1.92	1.16	1.96	1.22	14.36	9.76	14.89	10.95	3.92	2.71	4.08	2.89

This formulation has considerable appeal because it is readily extendable to other generalization situations, and because the mediating processes referred to are exclusively verbal. Thus, they are presumably subject to rather direct assessment and manipulation, and require no supplementary theoretical treatments of semantics or other philosophically redoubtable topics.

In spite of its appeal, this formulation of verbal generalization does not seem to account for the positive transfer found for words participating in the Type S interlist relation in Sequence I, where the required associative connections between lists were apparently lacking. However, this failure may not be in the formulation itself, but in the index of associative connection used in this experiment. The strength of some of the interlist associative connections between the words of Type S in Sequence I may have been considerably greater in the experimental situation than had been inferred from the word association test norms. That is, there may have been features of the experimental situation which produced a temporary increase in the strength of the interlist connections between some of the words of Type S sufficient to account for the limited positive transfer found.

A study reported by Storms (1958) provides evidence for the possibility that temporary supplemental strengthening of associative connections does occur under conditions similar to this. A list of words was read aloud to a group of *S*s who were told they were to be subsequently tested for their retention of these words. The list contained a number of words which appeared with very low frequency as associative responses to a set of stimulus words in an association test (median frequency was 1%). When the association test was administered under these conditions, the median net increase in responding with the words which had recently been heard was 18%. Evidently, these conditions made the usually weak response "more available" for association.

The present experiment provides a situation in which the strength of some of the associative connections between the words of Type S in Sequence I, although very low under most conditions, may have been temporarily supplemented in a manner similar to that found by Storms, for on the first exposure to the transfer list, *S*s were told to read the words aloud and try to anticipate them correctly thereafter. This possibility is empirically meaningful, for it may be possible to discover associative connections that are not susceptible to supplemental strengthening, or to present materials in generalization tests which do not provide the conditions for supplemental strengthening of weak associations.

SUMMARY

Transfer and recall of lists of paired associates were compared in situations in which the original and subsequent lists had identical stimulus words but in which the response words were variously related. The response words in the different lists were either strongly associated but not similar in meaning, highly similar in meaning but not associated, both similar in meaning and associated, or neither similar in meaning nor associated.

The results indicated that positive transfer occurred whenever the words were linked by associative connections, irrespective of semantic relations. Some positive transfer also occurred to a lesser extent across words highly similar in meaning but only weakly associated. It was suggested that the latter result was due to a temporary supplementation of some of the interlist associations connecting these words.

The results suggested a formulation of verbal generalization in which the tendency for the occurrence of the generalized responses is increased because of their associative connections to the responses learned in the original task. On the basis of this formulation, generalization between words judged to be similar in meaning occurs only because associative tendencies connecting them are usually present in *S*s' verbal behavior.

REFERENCES

COLLIER, R. O., JR. Analysis of variance for correlated observations. *Psychometrika*, 1958, **23**, 223–236.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

HAAGEN, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Unpublished doctoral dissertation, State University of Iowa, 1943.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

KOGAN, L. S. Analysis of variance: Repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.

MCCLELLAND, D. C., & HEATH, R. M. Retroactive inhibition as a function of degree of association between original and interpolated activities. *J. exp. Psychol.*, 1943, **33**, 420–426.

MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592–603.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 299–301.

OSGOOD, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, **38**, 132–154.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *ONR tech. Rep.*, 1954, No. 11. (Contract No. N8onr-66216, Univer. Minnesota)

RUSSELL, W. A., & JENKINS, J. J. Associative frequencies for Kent-Rosanoff stimuli and responses under various conditions. *ONR tech. Rep.*, 1956, No. 12. (Contract No. N8onr-66216, Univer. Minnesota)

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

SNEDECOR, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State Coll. Press, 1956.

STORMS, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, **55**, 390–395.

UNDERWOOD, B. J. *Experimental psychology: An introduction.* New York: Appleton-Century-Crofts, 1949.

UNDERWOOD, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.

(Received May 11, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 80–87

THE INFLUENCE OF THIRST AND SCHEDULES OF REINFORCEMENT-NONREINFORCEMENT RATIOS UPON BRIGHTNESS DISCRIMINATION[1]

ROY LACHMAN

Hollins College

Experimental variations of appetitional drive states have produced assorted behavioral effects. The results of research directed toward uncovering the relationship between drive and discrimination processes have been most conflicting (Birch, 1945; Dodson, 1918; Meyer, 1951; Myers, 1952; Tolman & Gleitman, 1949). This may be due in part to certain effects of deprivation time (T_d) that interact with other variables to produce the behavior observed.

Limiting our inquiry to brightness discrimination tasks, it would appear that various drive levels produce specific differences in limited aspects of the response. For example, an increase in T_d will result in, among other things, an increased rate of locomotion. Consider now the effect this will have upon other variables: faster running results in less delay of reinforcement (T_g); also faster running permits the stimuli to impinge on the organism's receptors for a shorter period of time (T_{sd}). Thus, a low-drive producing slow locomotion will result in a longer delay of reinforcement which may in turn slow up the acquisition rate for the choice behavior. In contrast, the longer latencies of low motivation will increase the stimulus duration time (T_{sd}) possibly resulting in quicker learning (Amsel, 1952). The relationship between drive level and $T_{sd} - T_g$ was one of the considerations that generated a "miniature theory," the details of which are presented elsewhere (Lachman, 1960).

The present study seeks to apply the delay of reward, stimulus duration interpretation of drive to the learning of a difficult discrimination. The superiority of high-drive *S*s should be limited to the learning of easy discrimination problems where a shorter delay of reward is produced by high-drive locomotion rates. In a difficult discrimination, the duration of choice-point stimulation is assumed to be critical. Here, the superior delay of reward for high-drive and the superior stimulus duration interval of low-drive might cancel each other. Drive level, therefore, would produce *no* differences in rate of learning. Differences in learning rate between high- and low-drive groups in a difficult discrimination should be reinstated if the choice-point T_{sd} level is equated for the respective drive groups.

Control of choice-point T_{sd} might be accomplished in a Y maze by employing transparent doors and a starting interval of sufficient duration. With transparent doors, gross differences might be eliminated in the duration during which high- and low-drive groups observe the cue characteristics of the choice point. A facilitative effect of high-drive should occur. If opaque doors are employed in a discrimination problem of the same high difficulty level, differences in learning rate due to drive should not occur.

---

[1] This article is based upon a dissertation submitted to the Department of Psychology, New York University, in partial fulfillment of the requirements for the PhD degree. The author is greatly indebted to his sponsor, H. H. Kendler, for his advice and encouragement throughout the course of this investigation. Thanks are also due M. R. D'Amato for technical assistance. The study was conducted during the author's tenure as a Samuel S. Fels Fellow.

Comparisons of discrimination attainment between conditions of opaque and transparent doors might be most interesting if forced choice training is employed. The two types of maze door vary the character of choice-point stimulation on forced choice trials and an empirical interpretation might be provided for the forced choice training procedure.

A problem of no minor theoretical importance is the relative influence of trials terminating with the selection of the reinforced and nonreinforced discriminanda alternatives. In the free response situation, the ratio of occurrences of correct and incorrect responses is an uncontrolled consequence of *S*'s behavior.

During discrimination acquisition with a nonpreferred S+ (e.g., the reinforcement of responses to the brighter stimulus for negatively phototropic organisms), the proportion of R+ responses begins in the neighborhood of .25 and terminates at unity. The initial R+ proportion for a preferred S+ is of the order of .75. Differences in acquisition with preferred and nonpreferred discriminanda may involve different learning rates (slopes) as well as unequal initial habit levels. Stated differently, the problem involves the relative influence upon discrimination acquisition of responses to S+ as compared with responses to S−. This problem may be investigated by keeping the initial habit level constant (employing the nonpreferred cue as S+) and varying the response ratio (R+ to S+)/(R− to S−) by forced choice trials.

The investigation thus concerns three areas of influence upon discrimination learning: drive level, choice-point cue availability, and the ratio of responses to the stimulus alternatives.

METHOD

Subjects.—The *S*s were experimentally naive, male albino rats obtained from the Blue Spruce Farms. At the start of the experiment, they ranged in age from 65 to 85 days. The results are based on data collected from 72 *S*s, 6 in a cell. Of the 36 low-drive *S*s, 4 were replaced owing to failure to run.

Apparatus.—The discrimination equipment[2] consisted of completely automatic, single-unit Y mazes, each possessing three identical arms. This arrangement permitted automatic programing of training conditions and dispensed with the handling of *S*s between successive trials of the same day. The four Y mazes were completely independent and identical in structure, except for opaque vs. transparent doors. Self-contained, electronic programing devices were available for each set of apparatus. The *S*s were run two or three at a time. The latency and choice data as well as the experimental program for each *S* on every trial were stored in "memory" equipment. Every 5 sec. a scanning device collected the data for each apparatus and transmitted it to a Clary "print-out" which recorded the information on a continuous tape.

At the start of the daily training session, *S* was placed in one of the three arms which served, for that trial, as a starting compartment. One of the remaining two arms contained S+, the positive stimulus, while the other one contained S−. After *S* made his choice and completed the trial, the arm which he chose became the starting compartment for the next trial. The other two arms contained S+ and S−. This procedure continued until *S* completed the number of trials programed.

Maintenance (experimental operations for the drive concept).—Two levels of water deprivation were utilized. High-drive *S*s were watered for 4 hr. on alternate days, 15 min. after being removed from the apparatus. Low-drive *S*s were watered for 10 min., 4 hr. prior to training, and were administered a second water ration for 1 hr., 15 min. after training. Since training took place every second day, on nontraining days, low-drive *S*s were watered for 10 min. followed by an additional 1 hr. of water, administered 5 hr. later. Food was present at all times in the home cage. A temperature of 78° ± 3° and relative humidity varying between 25% and 45% were maintained in the living quarters and experimental rooms.

---

[2] The apparatus was designed by M. R. D'Amato (cf. D'Amato & Jagoda, 1960).

Discrimination training.—A free choice or forced response to the bright arm (S+) was always rewarded with water; responses to the dark arm (S−) were never rewarded. The starting stimulus was a flashing light less than .2 ft-c in intensity; S+ was a continuous light of 1.6 ft-c.; S− .4 ft-c. The intervals employed were: stimulus, 10 sec. ± .5 sec.; intertrial, 25 sec. ± 2.5 sec. (lights out); starting, 8 sec. ± .5 sec. The water reinforcement duration was 2 sec. ± .5 sec., i.e. *S* was permitted to drink for an average of 2 sec. after making physical contact with the water.

Forced trials were employed to produce ratios of correct responses to incorrect responses of 3/1, 1/1, and 1/3. This was accomplished by programing blocks of 4 trials, 1 free and 3 forced. One of the forced trials was programed as contingent upon the free trial; the contingent trial always produced a response which was opposite to the free trial response. The remaining two forced trials were either both positive, or both negative, or one positive and one negative; the ratios stated were thus produced. Since responses to the brighter stimulus received 100% reinforcement and those to the darker stimulus received 0% reinforcement, *S*s with a correct/incorrect ratio of 3S+/1S− ended each interval of 4 trials with 75% of their training trials to the reinforced arm (Group 75%). Likewise, for each block of 4 trials, the 1S+/1S− and 1S+/3S− ratio groups were administered forced choices that resulted in 50% (Group 50%) and 25% (Group 25%) of their trials, respectively, being run to the correct stimulus. On the final day of training, all 40 trials were free. The final test trials were employed because postremity effects were suspected to be operating in the forced trial phase of training. That is, the response on a free trial may have been influenced by the direction of forcing on the immediately preceding forced trial. Furthermore, the free trials of the final day eliminated the long latencies typically encountered during the terminal stages of forced choice training (Spence, Goodrich, & Ross, 1959).

TABLE 1

ANALYSIS OF VARIANCE FOR THE NUMBER OF CORRECT RESPONSES ON DAY 5

Source	*df*	*MS*	*F*
Schedules (S)	2	108.10	3.50*
Drive (Dr)	1	36.13	1.17
Doors (Do)	1	342.35	11.08**
S × Dr	2	35.78	1.16
S × Do	2	1.26	<1
Do × Dr	1	1.12	<1
S × Dr × Do	2	34.13	1.10
Within	60	30.89	
Total	71		

* $P < .05$.
** $P < .01$.

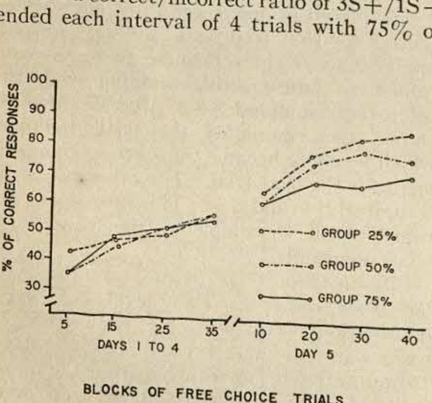


FIG. 1. The percentage of correct responses on free choice trials for forced choice response schedule groups.

All *S*s were given 20 trials on Day 1 and 40 trials a day on Days 2 to 5. A detailed account of forcing procedure and preliminary training will be found in Lachman (1960, Exp. 2).

Experimental design.—A 2 × 2 × 3 design was adopted: High (H) and low (L) thirst drive; transparent (T) and opaque (O) maze doors; and three ratios of response, 3/1 (75% choice of S+), 1/1 (50% choice of S+), and 1/3 (25% choice of S+).

RESULTS

Figure 1 shows the learning curves for the three response schedule groups. The curves begin at below-chance level because of *S*s' preference for S− (dark). By Day 4, differences are still absent and all groups are responding slightly above the chance level. The exclusively free choice trials of Day 5 exhibit divergent curves. During the terminal 10 trials of Day 5, Group 25% is responding close to the 90% correct response level, Group 75% is performing near the 70% level, and Group 50% is between the two.

Table 1 reports the analysis of variance for the number of correct responses on Day 5; Table 2 reports the individual cell means. The analysis produced two significant effects: schedules of forced choice responses and transparent-opaque doors. The average number of correct responses on Day 5 is 31.42 for Group 25% (i.e., the group trained with 25% of R+ and 75% of R−), 28.83 for Group 50%, and 27.21 for Group 75%. For transparent and opaque doors the correct response averages are, respectively, 31.33 and 26.97.

TABLE 2

Cell Means: The Average Number of Correct Responses on Day 5

Forced Response Schedules	High Drive		Low Drive	
	Transparent	Opaque	Transparent	Opaque
25%	33.5	28.8	34.0	29.3
50%	31.8	30.0	29.7	23.3
75%	31.2	23.8	27.8	26.0

The differences in learning rate for drive levels are not significant. The high-drive *S*s made an average of 29.86 correct responses on Day 5, the low-drive *S*s 28.44. No significant interactions were obtained.

Figure 2 graphically reveals the reason for the significant door effect. Group T (transparent doors) started and remained at a higher level of response throughout the training days of intermixed free and forced choice trials. On the totally free choice trials of Day 5, the performance of Group T continues to improve with terminal responses at the 87% correct-response level. By the second block of 10 trials for the same day, Group O levels off at the 70% level and remains there until the termination of training.

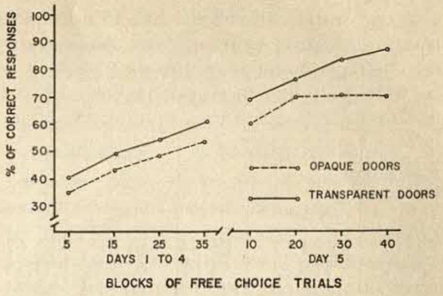


Fig. 2 The percentage of correct responses on free choice trials for transparent and opaque conditions.

Thus the choice data show that transparent doors in a combined free and forced choice situation facilitate discrimination learning and that the speed of this learning tends to increase with greater proportions of negative instances (incorrect, unrewarded, forced choices).

The results for the running data are presented in Fig. 3. The percentage of trials completed within

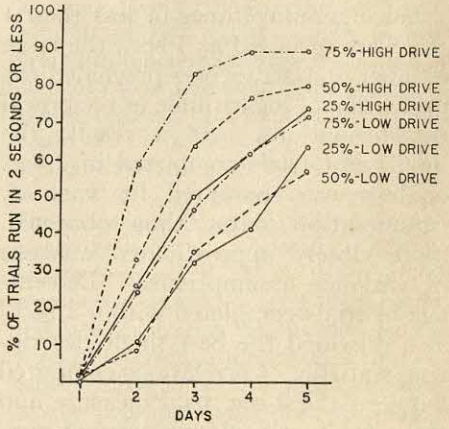


Fig. 3. The percentage of free and forced trials (Days 1 to 4) and free trials (Day 5) run in 2 sec. or less for drive and response schedule groups.

TABLE 3

CELL MEANS: THE AVERAGE NUMBER OF TRIALS COMPLETED WITHIN 2 SEC. FOR ALL TRAINING DAYS

Forced Response Schedules	High Drive		Low Drive	
	Transparent	Opaque	Transparent	Opaque
25%	84.17	85.83	67.67	51.67
50%	110.83	94.67	62.33	57.33
75%	136.00	122.17	80.17	86.33

2 sec. is plotted as a function of training days. Forced choice response groups for each of the drive levels are plotted as parameters. Each curve represents the running data for 12 *S*s since transparent-opaque cells were collapsed for graphic presentation. Figure 3 indicates that rate of running increases with the proportion of forced choices to the rewarded stimulus and with drive level. Group H-75% improves its locomotion rate in a highly accelerated fashion. The running time of Group H-50% is likewise facilitated but to a lesser degree. Group L-75% is facilitated in running speed to a degree that eliminates running differences between it and Group H-25%.

Since running times of less than 1 sec. were recorded as 1 sec., the large number of 1-sec. scores precluded the utilization of logarithmic or reciprocal transformations. As a result, the number of trials completed in 2 sec. or less was computed for each *S*. Enumeration data, thus obtained, more closely approximated analysis of variance assumptions. "Percentage of trials completed within 2 sec." also provided the best single descriptive statistic. Correlations computed between the 2-sec. trial measure and more conventional measures of running time were higher than the intercorrelations of the conventional locomotion scores (Lachman, 1960, Exp. 1.)

The running time averages for each cell is presented in Table 3. The analysis of variance based on these data is presented in Table 4. The score for each *S* is based on the total number of its trials completed with a running rate of 2 sec. or less. Forced choice schedules and drive level are both highly significant. No interactions are significant.

DISCUSSION

The results here obtained severely strain the credibility of the stimulus duration, delay of reward, drive model. Most damaging for the model are the locomotion data. Although both high drive and greater proportions of forced choice trials to the rewarded stimulus markedly accelerate locomotive behavior, only the latter influences choice behavior. If differences in delay of reward and choice-point stimulus duration are functionally related to discrimination acquisition, then acquisition effects should occur regardless of the ultimate source of the differential reward delay and cue duration. Response ratio schedules and thirst drive have similar effects on running rate and presumably upon the delay and duration variables; similarities in the acquisition of choice behavior should result. Clearly, this is not the case. Forced choice response schedules influenced both running time and discrimination learning. An increasing propor-

TABLE 4

ANALYSIS OF VARIANCE FOR THE NUMBER OF TRIALS COMPLETED WITHIN 2 SEC.

Source	*df*	*MS*	*F*
Schedules (S)	2	7,374.85	11.40***
Drive (Dr)	1	26,030.01	40.24***
Doors (Do)	1	931.68	1.44
S × Dr	2	737.93	1.14
S × Do	2	68.34	<1
Dr × Do	1	91.13	<1
S × Dr × Do	2	582.04	<1
Within	60	646.87	
Total	71		

*** $P < .001$.

tion of positive instances (rewarded forced choices of S+) increased the acquisition of the running response and perhaps its asymptote while simultaneously retarding the acquisition of the discrimination habit. In contrast, thirst level produced no influence upon the measure of discrimination learning, although it significantly influenced locomotion rate. These findings support the conclusion that discrimination learning is, to a degree, independent of the concurrent locomotive behavior and a concurrent or multiple habit hypothesis may be warranted.

Asymmetry of forced responses to the rewarded and negative stimuli poses an interesting problem. Group 75% ran three-fourths of its training trials to S+. Thus, this group received both a greater number of trials *and* a greater number of reinforced trials to the positive stimulus as compared with Group 25%. If a simple reinforcement model assumes no influence of negative (nonrewarded) instances, then according to either a reinforcement or a contiguity hypothesis Group 75% should have exhibited superior discrimination learning. The opposite was observed. It follows that *neither* a simple reinforcement nor contiguity theory can account for this result. The attempt to reduce the acquisition rules for discrimination learning to a single learning principle would appear to involve an oversimplification. Special rules of acquisition appear to be necessary in order to account for the choice data of discrimination learning. In contrast, a simple reinforcement assumption adequately describes the acquisition of the running response.

The present observations are consistent with the results of Denny and Dunham (1951) with rats and Cantor and Spiker (1954) with preschool children. These two investigations, unlike the present study, kept the number of positive instances constant and varied the number of negative instances.

The possible aversive effect of negative instances is by no means a novel formulation and has been considered, among others, by Spence (1958, p. 101). He postulated fractional antedating frustration reactions (r_f) for nonrewarded choice responses. Perhaps an r_g—r_f model is required to account for the present data. According to an r_g—r_f theory, the response schedule data (superiority of negative instances) would be interpreted as demonstrating a steeper acquisition gradient for r_f as compared with the acquisition of r_g. That is, classical conditioning of an avoidance response to the nonrewarded stimulus would appear to occur at a faster rate than the conditioning of r_g to the positive stimulus. If the correct choice response were assumed to increase in strength with higher levels of both r_g and r_f, the superiority of Group 25% would reflect the greater degree of conditioning of r_f.

It will be recalled that forced choice training was terminated *prior* to the appearance of difference in locomotion rate to S+ and S−. On Day 5, all trials were free choice. The locomotion rates on Day 5 for the several experimental groups maintain the same relative order as on the previous days when forced choice training was administered. Consequently, the graphic presentation and analysis of locomotion rates was based on data for all trials: free, correct forced choice, and incorrect forced choice.

Continuing the analysis according to Spence's theory, the superior locomotion rate of the 75% groups might be interpreted as reflecting a higher quantitative level of the intervening variable K. The slower running rate for the 25% groups could be interpreted as reflecting a greater level of I or a smaller relative level of K. In either case, the Hull-Spence assumption that "locomotion and choice behavior are equivalent indices of reaction potential" is in trouble. The results for the 75% and 25% groups indicate an inverse relationship between level of discrimination learning and locomotion rate under the initial conditions of the present study. Group 75% exhibited a greater rate of locomotion and an inferior level of discrimination learning. The reverse was observed for Group 25%. Actually, separate mathematical expressions are required to represent accurately the acquisition of choice and of running.

If we turn for a moment to the thirst data, a more serious difficulty arises for the hypothesis of locomotion and choice behavior as equivalent indices of reaction potential. In addition, the thirst data require a modification of the Hull-Spence assumptions concerning the interrelations between the intervening variables. Drive level produced a highly significant effect upon locomotion while at the same time producing no observable influence upon choice behavior. It follows that although the argument, $E+ = D \times H+$, is a satisfactory hypothesis for the running data, the expression, $(E+)$-$(E-) = (H+)$-$(H-)$, is more consistent with the choice behavior of the present research. There is an alternative manner by which Hull-Spence theory may accommodate the present effects of thirst. The assumptions concerning the coordinating definitions of D could be modified. That is, hunger and thirst manipulations may be assumed to have *different* influences upon D. This possibility warrants serious consideration.

The results produced by differential accessibility of choice-point cues resulting from the opaque and transparent doors require some explanation. Discrimination learning was significantly facilitated by the transparent doors. The transparent condition permitted additional reception of choice-point cues during the starting interval and observation of stimuli from the unavailable arm during the running of a forced trial. Thus, the enhancement effect may have occurred at either locus. A preliminary study (Lachman, 1960, Exp. 1) involved transparent doors without forced trials. Since the transparent doors did not enhance the discrimination in this preliminary study but did so in the present study, the locus of this effect appears to be at the choice point during forced trials. The argument, $H = f(N)$, might be refined so that N is defined not solely as the number of training trials but includes a term for the proportion of stimulus information available during training.

The discussion, thus far, has concerned theoretical approaches that include the noncorrection training procedure within their boundary conditions. Noncorrection maze learning has generated considerable difficulty for Estes' theory (1959, p. 416) especially in relation to the effects of nonreinforcement. The present results suggest that at least early in acquisition, nonreinforced responses have important consequences. A greater relative increment in the probability of R+ occurs after a response to S− than after a response to S+.

With respect to the model that initiated this research, it must be concluded that the model cannot be evaluated as true or false (or more appropriately as promising or unpromising). It is possible that the model is valid but the technique (thirst) used to evaluate it is not appropriate. The present data potently imply that the model may be invalid. To clearly demonstrate either of these two possibilities, more research is needed, perhaps with hunger instead of thirst.

SUMMARY

A model was proposed to explain the role of drive in discrimination learning. An experiment designed to test the model was conducted with three schedules of forced choice responses. That is, discrimination training was conducted with three different ratios of forced responses to the rewarded and nonrewarded stimuli.

The results and conclusions follow: (*a*) The research failed to produce the outcomes predicted by the model. (*b*) Discrimination learning was decreased by a relatively large proportion of forced choices to the *rewarded* cue and enhanced by a greater proportion of forced choices to the *nonrewarded* cue. (*c*) Level of thirst failed to influence the choice behavior in discrimination learning. However, thirst level exhibited a sizable influence upon running behavior. (*d*) Discrimination learning was improved by the presence of *both* relevant cues at the choice point during forced trials.

Several alternative interpretations of the results were explored, including the implica-

tions for the learning theories of Spence and of Estes.

REFERENCES

AMSEL, A. Rate of learning of a visual brightness discrimination as a function of discriminanda duration. *J. comp. physiol. Psychol.*, 1952, **45**, 341–346.

BIRCH, H. G. The role of motivational factors in insightful problem-solving. *J. comp. Psychol.*, 1945, **38**, 295–317.

CANTOR, G. N., & SPIKER, C. C. Effects of nonreinforced trials on discrimination learning in preschool children. *J. exp. Psychol.*, 1954, **47**, 256–258.

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials upon discrimination reversal. *J. exp. Psychol.*, 1960, **59**, 254–260.

DENNY, M. R., & DUNHAM, M. D. The effect of differential nonreinforcemrnt of the incorrect response on the learning of the correct response in the simple T maze. *J. exp. Psychol.*, 1951, **41**, 382–389.

DODSON, J. D. Relative values of reward and punishment in habit formation. *Psychobiology*, 1918, **1**, 231–276.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

LACHMAN, R. Drive level and forced choice response schedules in brightness discrimination. Unpublished doctoral dissertation, New York University, 1960.

MEYER, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.

MYERS, T. I. An experimental investigation of the effects of hunger drive upon the brightness discrimination learning of the rat. Unpublished doctoral dissertation, State University of Iowa, 1952.

SPENCE, K. W. Behavior theory and selective learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1958.* Lincoln: Univer. Nebraska Press, 1958.

SPENCE, K. W., GOODRICH, K., & ROSS, L. E. Performance in differential conditioning and discrimination learning as a function of hunger and relative response frequency. *J. exp. Psychol.*, 1959, **58**, 8–16.

TOLMAN, E. C., & GLEITMAN, H. Studies in spatial learning: VII. Place and response learning under different degrees of motivation. *J. exp. Psychol.*, 1949, **39**, 653–659.

(Received May 13, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 88–94

GENERALIZATION GRADIENTS IN A DISCRIMINATION SITUATION[1]

DAVID LaBERGE

University of Minnesota

The purpose of this study was to investigate the shapes of human generalization gradients in an effort to obtain information basic to a formulation of a set-model for stimulus generalization.

Gradients were obtained by first training *S*s to respond selectively to two different stimuli on a metathetic dimension (Stevens, 1957), and then testing them on new stimuli in the neighborhood of the training stimuli. The resulting gradients, plotted either in terms of relative frequency or latency, were expected to resemble one side of the traditional generalization gradient (Hull, 1939) which has a peak at the training stimulus and decreasing slopes on both sides of the peak.

The stimuli employed in the present experiments were positions of a small light along a vertical dimension. Since meaningful interpretations of shape characteristics of gradients required knowledge of the psychological spacing of the stimuli, preliminary scaling experiments were carried out to determine the optimal range of stimuli within the dimension to be used for generalization testing.

METHOD

Subjects.—The *S*s were men and women undergraduates from the University of Minnesota. The preliminary scaling experiment employed 16 *S*s and the generalization experiment employed 64 *S*s in the final analysis. All *S*s were naive with respect to experiments of this kind.

Apparatus.—The visual display was located 14 ft. in front of the *S*'s chair. The stimulus, a rectangle of light $\frac{15}{16}$ in. by $\frac{7}{16}$ in., could be positioned precisely along a vertical axis. In order to stabilize the perceived position of the light in the darkened room, two white enameled guide strips, $1\frac{1}{2}$ in. by 68 in., ran the length of the vertical axis on either side and adjacent to the rectangle of light. The lower border was 6 in. above the floor and the upper border was 76 in. above the floor. The position of a stimulus along the dimension was designated according to its distance in inches from the lower border. Thus Stimulus 35 was the physical midpoint of the range of stimuli bordered by the ends of the two white guide strips. Generous use of black commando cloth and flat black paint virtually eliminated all other visual reference points.

The *E* sat behind a light-proof screen and adjusted the position of the stimulus with the aid of a finely graduated scale. The rectangle of light was not illuminated during the time when *E* changed its position, and the room was sufficiently dark so that *S*s could see the rectangle only when it was illuminated.

A response box was located on a table in front of *S* and contained two levers, $5\frac{3}{8}$ in. apart, with red reinforcement lamps located $2\frac{5}{16}$ in. above each lever. An ATC timer[2] controlled the exposure times of the stimuli and reinforcements as well as all interevent intervals. A Beckman counter[3] measured response latencies in .001-sec. units, and a Beckman digital recorder printed out automatically the response latency, response alternative, and reinforcement alternative on each trial.

Scaling experiment.—The scaling procedure was based on measurement of difference thresholds. The simplest technique seemed to be to find a range of stimuli over which equal physical intervals would correspond to equal psychological intervals. The method of constant stimuli was used to obtain the

---

[1] This research was supported by a grant to the author from the National Science Foundation. Equipment was purchased through a grant from the Graduate School of the University of Minnesota.

[2] Automatic Timing & Controls, Inc., King of Prussia, Pennsylvania.

[3] Beckman Instruments, Inc., Richmond, California.

threshold curves around two standard stimuli, 41 and 45, which were near the border of a range of stimuli which previous studies had indicated as having equal and stable jnd's.

The procedure was to present a series of stimuli, the first of each pair being the standard, and the second a variable stimulus +2, +1, 0, −1, or −2 in. from the standard. The *S*s were instructed to determine if the second stimulus of each pair was above, below, or the same as the first stimulus. The pairs were randomized in 30-trial blocks, each block containing 6 of each of 5 possible standard-variable pairs. Each *O* was given 8 blocks, totaling 260 trials. The two standards were assigned to successive pairs of blocks in the orders 41, 45, 45, 41 or 45, 41, 41, 45. There were 1-min. rest periods between each pair of blocks. Five pairs of stimuli were given before each block to orient *S* to each new standard. There were four different random sequences, four *S*s to a sequence. The 16 *S*s were run two at a time, and were instructed to write their judgments on a tablet of paper. They were also instructed to look down at their hands immediately following the second stimulus of each pair in order to attenuate an autokinetic-type of effect found when *S*s focused rigidly on the spot where the last stimulus appeared.

The time relations of the constant stimulus experiment were as follows: A buzzer sounded for 1 sec.; time between the buzzer and the standard stimulus, 1 sec.; duration of the standard stimulus, 2 sec.; time between the standard stimulus and variable stimulus, 4 sec.; duration of the variable stimulus, 2 sec.; time between the variable stimulus and the buzzer of the next trial, 4 sec.

Since the psychophysical curves obtained for each of the two standards were virtually identical, we may conclude that within the range of Stimuli 41–45, and probably even within the greater range of Stimuli 39–47, equal physical intervals correspond to equal psychological intervals. Therefore we may reasonably expect to determine shape characteristics of generalization gradients obtained over this range of stimuli.

Procedure.—For all generalization conditions *S*s were given sufficient training trials with two stimuli, S_1 and S_2, to bring responding at or near asymptote. Then a series of generalization test trials to neighboring stimuli followed, with training stimuli interspersed among them to maintain discrimination performance at asymptote. Reinforcing lights always followed training stimuli such that one response, A_1, was reinforced 100% of the time when S_1 occurred, and the other response, A_2, was reinforced 100% of the time when S_2 occurred. Neither reinforcing light followed occurrences of test stimuli.

There were three discrimination training intervals: Stimuli 42–44, 41–45, and 27–59, which were labeled the 2-, 4-, and 32-in. training intervals. All training intervals had a common midpoint, namely Stimulus 43. Sixteen *S*s were assigned to each of the three discrimination conditions and were given generalization tests on seven stimuli lying between the training stimuli, S_1 and S_2. In addition, eight *S*s were assigned to the 2-in. and 4-in. training intervals and were given generalization tests to seven stimuli, six lying outside the training interval and one lying midway between the training stimuli.

The stimuli were randomized without restrictions within blocks of 17 trials. Each *training* block contained eight S_1, eight S_2, and one test stimulus, 43. Each *test* block contained five S_1, five S_2, and seven different test stimuli.

One *S* was tested at a time, and given 35 blocks of trials, totaling 595 trials. The blocks were given in five series, seven blocks to a series, with 3-min. rest periods between each series. The first series contained seven training blocks. Subsequent series contained an initial training block followed by six test blocks. A different sequence of stimuli was constructed for each subgroup of four *S*s. Side of the reinforcing light was counterbalanced against S_1 and S_2 within each sequence.

Time relations within a trial were as follows: Duration of the stimulus light, 2 sec.; time between stimulus and reinforcing light, 1 sec.; duration of the reinforcing light, 1 sec.; time between reinforcing light and next trial, 3 sec.

The *S*s were instructed to press a lever when a white light appeared in the front of the room. They were told that a red light above one of the levers would flash to indicate which lever they should have pressed on that trial. In addition, they were told that sometimes no reinforcing light would follow a white light and that such trials were to be considered as test trials wherein *E* wanted to know which lever *S* thought should be pressed on that trial.

RESULTS AND DISCUSSION

Discrimination training.—Discrimination curves for *S*s who learned the discrimination and were given generalization tests are shown in Fig. 1.

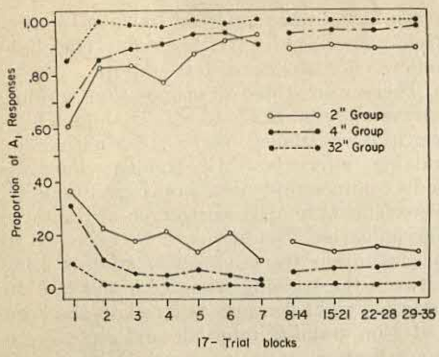


FIG. 1. Discrimination curves for *S*s over all trials, 16 *S*s in each group. (The upper and lower curves represent responses to training stimuli S_1 and S_2, respectively.)

The learning criterion was 13 or more correct responses in Block 7 of training. The data from the 2-in. and 4-in. groups which were given generalization tests to the outside are not included here, although their data were the same as the 2-in. and 4-in. groups which had tests to the inside of the training intervals. Data from training stimuli of training and test blocks were combined in Fig. 1, since none of the *t* tests of differences between test and training blocks approached significance. There are clear separations of the three discrimination groups, with an *F* test of the differences among asymptotes being significant beyond the .001 level. The curves of discrimination learning for all *S*s, learners and nonlearners, show even greater separations, and all groups appear to have reached their asymptotes by Blocks 5 or 6. There is no indication that the obtained orderings of group asymptotes will eventually reverse, although a reversal is predicted by a Markov model of discrimination learning proposed by Atkinson (1958). Test trials apparently do not reduce the level of discrimination performance as indicated by testing the differences between the last two blocks of the first series and the seven blocks of the final test series for all groups, including the groups having generalization tests to the outside of the training interval.

Generalization tests.—Since we were interested in generalization gradients

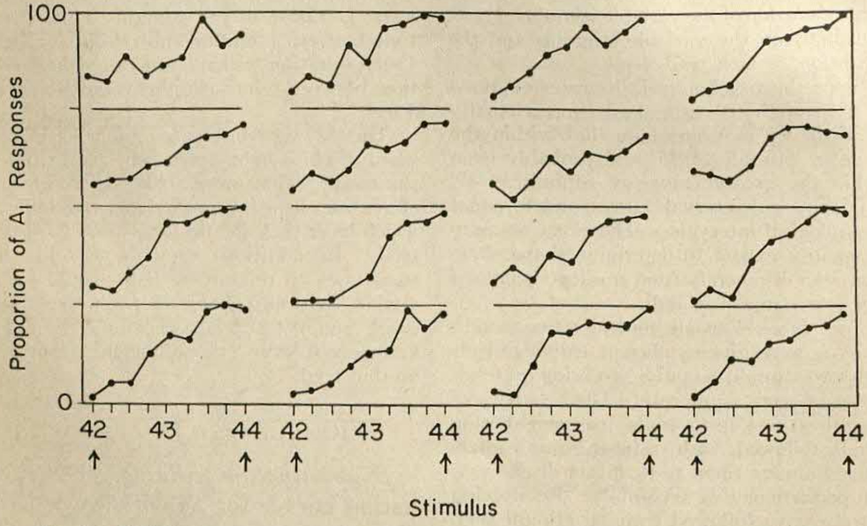


FIG. 2. Individual gradients for the 2-in. training group, where testing was made inside the training interval.

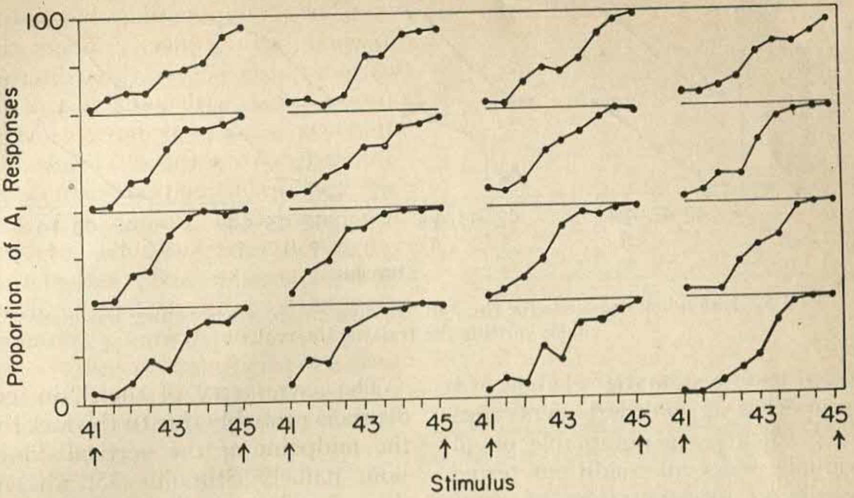


FIG. 3. Individual gradients for the 4-in. training group, where testing was made inside the training interval.

only of *S*s who learned the discrimination, we did not use generalization data from the seven *S*s of the 2-in. condition and five *S*s from the 4-in. condition who did not learn the discrimination in the first seven training blocks.

Individual gradients for each of the discrimination conditions are shown in Fig. 2, 3, 4, 5, and 6. For these curves, each training-stimulus point is based on 120 observations, and each test-stimulus point is based on 24 observations. The curves were obtained by averaging across all test blocks, a procedure justified by the observation that series by series comparisons of group generalization curves

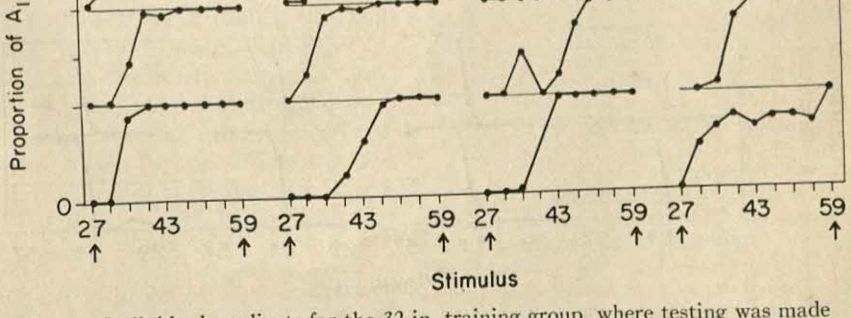


FIG. 4. Individual gradients for the 32-in. training group, where testing was made inside the training interval.

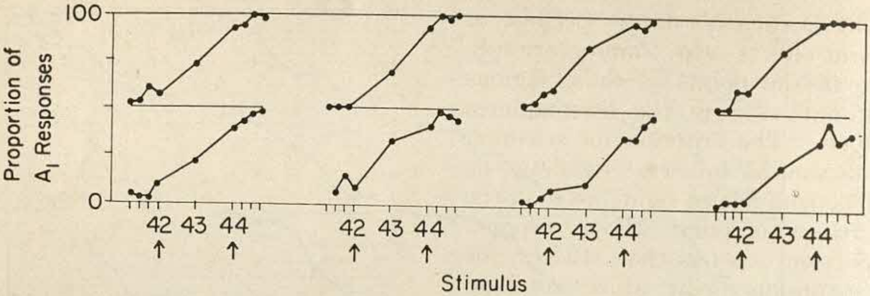


FIG. 5. Individual gradients for the 2-in. training group, where testing was made chiefly outside the training interval.

showed no systematic change. It appears that fairly orderly generalization gradients are obtainable on individuals under all conditions tested. Meaningful interpretations of shape characteristics of the gradients are justified only in the case of the 2-in. and 4-in. inside groups, where the test stimuli fall within the range of values scaled by the preliminary psychophysical experiment.

A notable feature of the individual gradients is their almost linear slopes, which might be expected in cases where stimuli are equally spaced. There is no consistent indication of plateaus near the middle of the gradients, even in the case of the 32-in. group. However, a strong case can be made for the existence of plateaus at either end of the gradients near the training stimuli for certain individuals who learned the discrimination to a high criterion.

The asymmetry of the 32-in. gradients is probably due to the fact that the midpoint of the vertical dimension, namely Stimulus 35, was not the midpoint of the 27–59 training interval. And verbalizations of *S*s confirmed that the dimension's physical midpoint served as a strong reference point for most judgments of the 32-in. condition.

In the cases of the generalization tests to the outside of the S_1-S_2 interval shown in Fig. 5 and 6, there is a shift in the peak response rate away from the training stimuli, an effect similar to that found in the pigeon studies of Guttman and Kalish (1956) and Hanson (1959).

The slope characteristics of the group gradients may be compared by referring to Fig. 7, 8, and 9. Statistical tests of differences in slopes were prepared by first computing the slopes of a least-squares straight

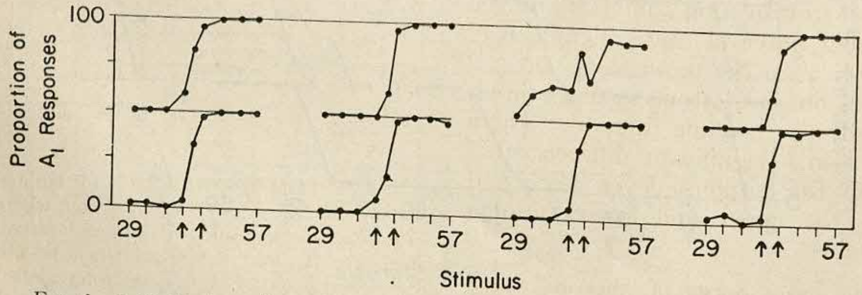


FIG. 6. Individual gradients for the 4-in. training group, where testing was made chiefly outside the training interval.

line fit to each *S*'s data. Because of plateau effects near training stimuli, some of the points in these regions were not used in the least-squares analysis. The criterion for selecting points was as follows. Reading the graphs from left to right, we took the last stimulus point whose response proportion was less than .10 and the first stimulus point whose response proportion was greater than .90, plus all stimulus points between these two points. The resulting distributions of individual slopes had little overlap and *F* tests of differences among slopes for the three discrimination intervals were highly significant. For both the 2-in. and 4-in. conditions, the mean slopes of the gradients obtained from testing stimuli inside the training interval were significantly greater than the mean slopes of gradients obtained from testing stimuli outside the training interval. There was also a significant difference between the response lever associated with the upper and lower training stimulus.

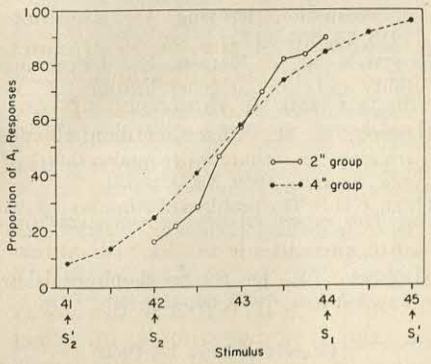


FIG. 7. Group frequency gradients for the 2-in. and 4-in. training groups, where testing was made inside the training intervals. (Each curve represents data of 16 *S*s.)

The exact form of the individual generalization gradient could not be determined from the present data, in spite of efforts to provide careful measurements of stimulus spacings for the 2-in. and 4-in. groups. The difficulty stemmed from the fact that all curves, including individual curves, were averages of more basic gradients. The sigmoid shape of the group curves of Fig. 7 can be obtained from averaging linear curves which have plateaus of varying length near the training stimuli. And similarly, an individual gradient would assume a sigmoid shape if fluctuations in acuity during the testing series produced plateaus of varying length. Until a method is found for obtaining generalization gradients under rigidly

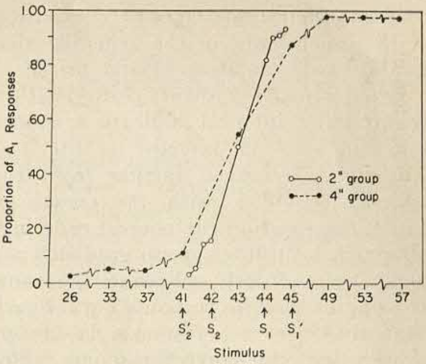


FIG. 8. Group frequency gradients for the 2-in. and 4-in. training groups, where testing was made chiefly outside the training interval. (Each curve represents data of 8 *S*s.)

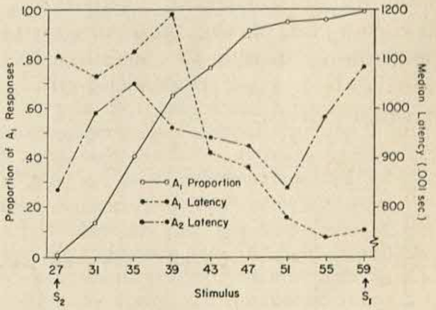


FIG. 9. Group frequency and latency gradients for the 32-in. training group, where testing was made inside the training interval. (Response A_1 was reinforced when S_1 appeared and response A_2 was reinforced when S_2 appeared. Each curve represents data of the same 16 *S*s.)

constant conditions, the determination of the exact shape of the generalization gradient will remain an elusive prize.

Group latency gradients plotted within the training interval indicate a monotonic increase in latency as the test stimulus is removed farther from the training stimulus, with the curves of Fig. 9 representing the poorest orderings obtained. All other group gradients are much more orderly, showing the same crossing of the two response curves such that the incorrect response is always slower than the correct response. For the cases where tests were made outside the training interval, the group curves continued their monotonic form over the range of outside test stimuli, with the fastest latencies being associated with the test stimuli outside the training stimuli.

SUMMARY

Sixty-four *S*s were given 119 trials on a discrimination problem involving two different positions of a light on a vertical dimension. Following this they were given 476 trials made up of (*a*) generalization tests to neighboring stimuli along the dimension; and (*b*) further training trials with reinforcements to maintain asymptotic discrimination performance during testing.

Individual and group frequency generalization gradients were fairly orderly. The magnitude of the interval between training stimuli had a highly significant effect upon the slope characteristic of the gradients. Peaks of the gradients were shifted toward the outside of the interval between training stimuli whenever response rates to the training stimuli were below maximum. Group latency gradients indicated similar monotonic orderings under all conditions.

Discrimination asymptotes were ordered in the expected way with no appreciable change in asymptotic level over the generalization test trials.

REFERENCES

ATKINSON, R. C. A Markov model for discrimination learning. *Psychometrika*, 1958, **23**, 309–322.

GUTTMAN, N., & KALISH, H. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

HANSON, H. M. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, **58**, 321–334.

HULL, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

(Received May 18, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 1, 95–96

SUPPLEMENTARY REPORT: EFFECT OF PRACTICE ON AN ILLUSION [1]

DOROTHY RETHLINGSHAFER AND THOMAS I. SHERRER

University of Florida — *Florida State Employment Service*

In a recent study, Rechtschaffen and Mednick (1955) found that *S*s in darkness would experience autokinetic movement that reflected the suggestion that a pinpoint of light would write words. Wide individual differences were noted, but no practice effects were studied. The present investigation is an attempt to determine the practice effect, if any, in the autokinetic illusion when *S*s are given the instructions that the light will write English letters and words.

Method.—The *S*s in the major experimental group were 40 male college students (age range = 17 to 28 yr). Nine men and two women from the same population were used in a control group.

Each *S* was tested on three 20-min. sessions on different days. During the first 10 min. of each session *S*s were told that the small pinpoint of light would write letters. A rest period, approximately 5 min., was given before a final 10 min. when *S*s were told that the light had been rearranged so that it would write words.

After first determining whether the illusion of movement was perceived, the first instruction was given:

> The first 10 min. part of the test will be essentially this: the light has been set up so it will trace the outlines of the letters of the alphabet. The letters will vary in size, and will be given in continuous series, with no breaks between them. As soon as you can make out the letters, or distinguish what they are, report them to me as you go along. There will be both capital and lower case letters given in manuscript form just as though I were writing them on a blackboard for you to read. Okay?

After the rest period, the following instructions were given:

> The second part of the test is essentially the same as the first, with the exception that the letters have been arranged so that words will be spelled. We aren't accustomed to using individual letters in reading, writing, or in speaking; for this reason it is believed that it will be easier for you to group the letters into familiar words, rather than to break the series into individual letters. There will be no breaks between words, so this is primarily a test to see if you can tell when one word is complete and another begins. If you make out some of the letters and miss some, you may be able to guess the whole word by the order in which the letters are given. As soon as you have completed the word tell me what it is. Okay?

The *S*s were seated in a dark room, approximately 12 ft. from the pinpoint of light. The *E* sat near the box from which the light was emitted and recorded responses. A luminous dial watch was used to record time.

The data reported were obtained from two groups. In Group 1, *S*s made their reports in Sessions 1 and 2 alone and in Session 3 with another *S*. In Group 2, *S*s were with another *S* in Sessions 1 and 2, and alone in Session 3.

Ten control *S*s were tested alone with the same instructions as used in Session 1, *but no pinpoint of light was present.* No control *S* reported perceiving letters or words. Five reported slight pinpoints or "flashes" of light which disappeared when attempts were made to focus upon them. The results are reported only for the experimental *S*s.

Results.—The experimental *S*s increased in the frequency of their reports throughout the three sessions. Practice affected the perceiving of the autokinetic illusion in the particular form induced by the suggestion given. As seen in Table 1 there was a greater frequency of responses in Session 2 as compared with Session 1. This was true for the reports of both letters and words, and for both groups. Except for reports of letters by Group 1, there was a further rise in frequency of responses in Session 3. Tests of significance of the mean differences indicated that the increases from Sessions 1 to 3 were statistically significant ($P < .05$) except for Group 1 in their reports of letters. For Group 1, the rise in frequency of reports of letters from Session 1 to Session 2 was significant ($P < .05$). The between-sessions F was significant at beyond the .01% level for both groups on words and letters except for the Group 1 words measure, which was significant at beyond the .05 level.

The number of responses of letters was 1005 with all letters of the alphabet being reported. The letter most frequently perceived was *e*, with *m* and *o* being next. There were 238 words perceived, most of them being fairly simple and short words as: *the*, *be*, *of*, *all*, etc. Almost half of the words were reported only once, with a few of these being individual, such as, *Robert*, *nonsense*, *bombs*, etc. Some *S*s accepted the directions readily. They also shifted easily from the set to perceive letters to perceiving words. At the other extremes were *S*s who questioned the instructions. They gave few responses and with qualifications.

Certain *S*s repeated the same responses; others experienced unusual "writing," such as perceiving the letter as being written on its side, or as movement from left to the right, making the writing difficult to perceive.

It is important to note that *S*s must accept the suggestion that the light is writing. When graduate students were used as *S*s in a preliminary study, the number of responses were few or nonexistent. The *S*s used in this report were naive in respect to the illusion but some were more critical than others of the suggestion that the light was writing.

In this study, the particular form of the illusion was induced in *S*s from the instructions given by *E*. As in the Atkinson and

---

[1] This study was reported, in part, at the September 1957 meeting of the American Psychological Association.

TABLE 1

Effect of Practice on an Illusion

Session	Letters				Words			
	Group 1		Group 2		Group 1		Group 2	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	8.10	5.46	5.35	3.29	1.55	2.01	1.00	1.35
2	11.55	9.59	7.15	2.85	1.70	1.82	2.25	1.93
3	9.95	7.87	8.15	4.05	2.70	2.28	2.70	2.19

Ammons (1952) study, practice seems to affect the stability of a perceptual organization when it is set up by external directions. However, it is possible that practice also affects perceptual organizations when some particular directional set is adopted by *S* himself, either by conscious self-instructions or by sets carried below his level of awareness. In this way, perceptual "habits" may be strengthened below the level of awareness. The ambiguous world, or those aspects of it which are most ambiguous, would thus by practice be perceived in an individual manner.

References

Atkinson, R. C., & Ammons, R. B. Experiential factors in visual form perception: II. Latency as a function of repetition. *J. exp. Psychol.*, 1952, **43**, 173–178.

Rechtschaffen, A., & Mednick, S. The autokinetic word technique. *J. abnorm. soc. Psychol.*, 1955, **51**, 346.

(Received August 6, 1960)

Journal of
Experimental Psychology

VOL. 62, No. 2 AUGUST 1961

PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF SIMILARITY: COMMON STIMULUS AND RESPONSE ITEMS WITHIN THE LIST

TAKAO UMEMOTO[1] AND ERNEST R. HILGARD

Stanford University

Although in serial learning the same item serves as both stimulus and response, when a paired-associate list is arranged in comparable fashion (e.g., A-B, B-C, C-D, etc.), learning is more difficult than in simple serial learning (Primoff, 1938). It has been found, further, that familiarity with the items as a result of prior serial learning facilitates the learning of a specially constructed paired-associate list in which the same items appear as stimuli and as responses, but this facilitation is lost when the list is half-mastered (Young, 1959). The interferences that account for the difficulties within such paired-associate lists have been attributed by Primoff and by Young to backward association, an interpretation coherent with the findings of Murdock (1956) and Feldman and Underwood (1957).

In the two experiments to be reported we manipulated the similarity between stimulus and response items in a manner to demonstrate circumstances under which the phenomena shown by Primoff and by Young are likely to be found, and some circumstances under which they will not be found. Such manipulation has been done in somewhat different contexts by Underwood (1953a, 1953b, 1954) and by Morikawa (1959b). While there are several possible approaches to the manipulation of S-R similarity in paired-associate learning, we have chosen to take the approach in which similarity between S and R is controlled by the number of items in common between S and R.

EXPERIMENT I

In Exp. I, paired-associate lists were constructed with three degrees of S-R similarity: complete similarity, in which all S items appear as R items; half similarity, in which half the S items are found as R items, the other S and R items all being different; complete dissimilarity, in which all S items and R items are unlike.

What kind of effect can be expected in such a situation? Two contradictory outcomes appear plausible:

1. The first hypothesis can be stated as follows: the more common items there are, the more competitive and conflicting

---

[1] Now at Kyoto University, Japan. The experiments were performed while a visiting scholar at Stanford University, 1959–60.

TABLE 1

Lists Used in Experiment I

List 1	List 2	List 3
BAF-JOC	BUF-RAQ	BEF-SUC
DEX-SIV	CAG-MUZ	CIM-ROX
JOC-WEY	DOX-FID	DAT-FEN
LIG-HOK	LIY-JOW	JUP-HIB
PUM-DEX	WER-CAG	KIR-JAD
HOK-BAF	SOK-BUF	WOV-ZEK
WEY-LIG	QAL-DOX	TOG-PIQ
SIV-PUM	MUZ-HEV	HAC-LUV

responses will occur, hence more trials will be required for mastery. When *S* tries to learn a pair A-B in a list, and there is another pair C-A which has a common item as its response, a backward association, that is the association A-C, will also be acquired. Hence when the stimulus item A appears, there might be a competition between A-B and A-C, thus increasing the difficulty of learning.

2. The second possible hypothesis is: The ease of learning a list is inversely proportional to the number of different items in the list. Thus the list with more common items will have fewer different items to be acquired, and may lead to more rapid learning.

The first experiment tests these hypotheses, and should permit a choice between them. They are not mutually exclusive, but one effect should be stronger than the other, even if both effects are found.

Method

Lists.—Three experimental lists and one practice list were used. Each experimental list consisted of eight pairs of nonsense syllables, having a Glaze (1928) association value of 73%. The three lists are given in Table 1.

In List 1, each of the eight syllables is used both as S and R. In List 2, half of the eight syllables used are in common to S and R, so that there are 12 different syllables in this list. In List 3, none of the eight syllables is repeated, so all 16 syllables are different. In these lists similarities within the S and R items were equalized as well as possible. Some repetition of vowels is inevitable but each consonant and vowel appears no more than twice in the stimulus or the response side of these lists. In pairing syllables also an attempt was made to avoid any apparent connection between them.[2] The practice lists consisted of three pairs of nonsense syllables, with association values of 100% in Glaze's table.

Conditions and Ss.—Three groups of eight *S*s served in this experiment. All *S*s were students in introductory psychology at Stanford University, and all were naive with respect to verbal learning in the Laboratory.

After general instructions to memorize, *S*s were told:

> . . . pairs of nonsense syllables will be presented in the window of the memory drum, such as JIN-TAK, etc. The presentation time for each pair is 2 sec. Then the recall trial will begin in which the stimulus syllable on the left side is presented, such as JIN. You must recall the other item which was on the right side. . . . The order in which the syllables are presented varies from trial to trial. . . . Please *spell out loud* the correct answer (only) in the recall trial, e.g., TAK. Since the time allowed for recall is only 2 sec. please answer as quickly as possible. If you cannot remember the syllable, you may guess or keep silent, just as you wish. The learning trials and recall trials will alternate until you are able to remember *all* the items. There are two kinds of lists. The first is for your practice and composed of only three pairs. The second list is the main list, which is composed of eight pairs. Do not be in too much of a hurry to succeed. University students usually need 40 or more trials to memorize a list of eight nonsense pairs.

Lists were presented on a memory drum at a 2-sec. rate, with a 4-sec. intertrial interval. The learning trials and recall trials were alternated. The serial orders of the stimulus syllables were changed in each learning trial and in the recall trials in order to minimize serial learning.

The *S* learned the practice list to one errorless trial, then learned the experimental list to one errorless trial, after which *S* was asked to make a backward recall. The response syllables typed on cards were presented to *S*; he was asked to write down each stimulus syllable previously attached to the response syllable. Ten seconds were allowed to recall each stimulus syllable in this test. The *S*s were not informed beforehand about the later backward recall.

---

[2] There is one unintentional peculiarity in the construction of List 2 that may be noted. While half the S items recur also as R items, in only one pair (CAG-MUZ) are both members of a pair repeated in other parts of the list. This produces an unfortunate asymmetry in design. Because this relation is universal in List 1 (all items appearing both as S and as R) any contrast between results in Lists 1 and 2 cannot be attributed to the CAG-MUZ pair in List 2.

Results

The mean numbers of trials required to memorize the three pairs in the practice list were 3.38, 3.50, and 4.13 for Groups 1, 2, and 3, respectively. The differences between the scores are not statistically significant ($F = 0.46$, $df = 2/23$).

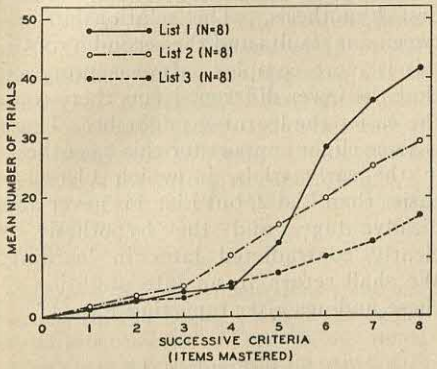


FIG. 1. Mean number of trials to successive criteria (Exp. I). (List 1 has all items common to both S and R, List 2 has half of the items in common, List 3 has none in common.)

The mean numbers of trials to reach successive criteria on the three experimental lists are shown in Fig. 1. In terms of the number of trials required to reach 8/8 correct responses, List 1 is most difficult and List 3 is easiest ($F = 19.25$, $df = 2/23$, $P = .01$). However, Fig. 1 also shows that in the early stage List 2 is rather more difficult than List 1. After the stage of 50% learning (4/8 criterion) the success with List 1 is delayed as shown by the rising curve. The learning process is somehow interrupted when the list is half learned, in agreement with Young (1959). This disruption is evident also when the average number of errors per list is similarly plotted. Table 2 shows the cumulative errors in the first half stage and later half stage. An analysis of variance based on this error score is shown in Table 3. The differences are significant not only with reference to List and Stage, but also in the interaction of these two variables.

TABLE 2

MEAN CUMULATIVE ERRORS IN EARLY AND LATE STAGES OF LEARNING

List	*N*	Stage I (0–50% Learning)	Stage II (50–100% Learning)
1	8	27.75	141.25
2	8	57.63	62.50
3	8	28.50	30.13

The individual data for List 1 show that learning develops smoothly at the first stage and *S*s easily reach the criteria of 4 or 5 out of 8 items. Thereafter almost every *S* suddenly shows confusion. They need a number of trials to acquire even one more correct response. This is also evident in the types of errors made (Table 4). The *S*s who learned List 1 not only show an extraordinarily high proportion of overt errors (51% vs. 23%

TABLE 3

ANALYSIS OF VARIANCE OF ERRORS FOR STAGES OF LEARNING (EXP. I)

Source	*df*	*MS*	*F*
Between lists (L)	2	12,288.9	12.62*
Between *S*s in same group	21	973.9	—
Total between *S*s	23		
Between stages (S)	1	19,280.1	31.55*
S × L	2	16,340.9	26.73*
Pooled *S*s × S	21	611.2	—
Total within *S*s	24		
Total	47		

* $P < .01$.

TABLE 4

ERRORS CLASSIFIED BY TYPE

List	*N*	Overt Errors		No Response	Total Errors
		Backward Associations	Other		
1	8	51.2	34.0	83.8	169.0
2	8	3.6	20.8	95.7	120.1
3	8	0.0	12.0	46.6	58.6

and 20% for Lists 2 and 3) but also show a predominance of errors classified as backward associations on the basis of possible mediation by common items. The difficulty experienced by *S*s learning List 1 in overcoming their trouble in the middle of the process was striking. To fill the most difficult gap (between the fourth and fifth, or fifth and sixth item recalled) took at least 15 trials for every one of these 8 *S*s. Only one of the remaining *S*s took that long. This effort to gain one item in recall required for List 1 *S*s more than a third of their total trials to mastery.

These deadlocks with List 1 almost always originated in perseverative repetition of wrong responses, which were mediated by common items in S and R. If there were A-B and B-C pairs in a list, and if *S*s learned A-B at first, it happened quite often that they responded B-A instead of B-C when B was later presented. Such a backward association mediated by common items in S and R is probably one of the interference sources in this situation, as proposed in our first hypothesis.

The mean numbers of correct responses in backward recall were for List 1, 5.38, for List 2, 5.75, and for List 3, 3.13. The prevalence of backward recalls in Lists 1 and 2 suffices to make our hypothesis plausible. The fact that there were about an equal number for the two lists is immaterial, for the opportunity to interfere is twice as great for List 1 as for List 2. The relatively lower number of backward recalls for List 3 calls for a comment. The backward associations tend to develop with fully practiced responses, as proposed by Umemoto (1951) and Morikawa (1959a). Because there was much less overlearning of the items of List 3 than of the items of List 1 and 2, the lower number of backward associations within List 3 is understandable.

The results thus give support to our first hypothesis. The relationship between our results and the second hypothesis is more complex. It was proposed that the fewer different items there were the easier the learning might be. There is some slight support for this hypothesis in the early trials, in which List 1 is easier than List 2, but List 3 is never at a disadvantage, and the hypothesis is clearly contradicted later in learning. We shall return to an interpretation of these findings after reporting Exp. II.

EXPERIMENT II

The results of Exp. I have been tentatively explained by the competition between right and wrong associations in Lists 1 and 2 which seem to have been mediated by backward associations. However, mediated backward associations need not always produce inhibition. If, for example, the mediated responses are the same as the correct response these mediated responses may facilitate

TABLE 5

LISTS USED IN EXPERIMENT II

List 1	List 2	List 3
YAV-perfect	YAV-perfect	YAV-perfect
noonday-GID	noonday-GID	noonday-GID
ZOT-clumsy	ZOT-vacant	ZOT-clumsy
empty-KEB	empty-KEB	empty-KEB
GID-midday	GID-midday	GID-vacant
faultless-YAV	faultless-YAV	awkward-YAV
KEB-vacant	KEB-clumsy	KEB-midday
awkward-ZOT	awkward-ZOT	faultless-ZOT

learning instead of interfering with it. When the mediated responses are quite different from the correct ones, we would expect competition, as in Exp. I. Experiment I controlled the possible frequency of mediated responses by adjusting the number of common items between S and R, and all the mediated responses (if any occurred) were incorrect. In Exp. II the possible frequency of mediated responses remained the same for all three lists, but the relation of the mediated responses to the correct response differed; that is, the items of the three lists were all alike except that the lists were so constructed that some backward associations would facilitate, some interfere. Thus any differences in results are unlikely to be attributable to the *number* of mediated backward associations, although they may be due to the *arrangement* whereby the mediated response is inhibitory or facilitating.

Method

Lists.—Three experimental lists and one practice list were used. Each of the experimental lists consisted of eight pairs of items, in every case a nonsense syllable paired with an adjective. Four out of eight pairs in each list had nonsense syllables on the stimulus side and adjectives on the response side; the remainder were in the reverse order. The nonsense syllables were all of low association value (Glaze 13%). The adjective pairs were taken from the Melton and Safier table (Hilgard, 1951), all from the 10 pairs highest in similarity. The three lists are given in Table 5.

As shown in Table 5, the three lists all have the same components, but differently paired. In List 1, each member of the four pairs of similar adjectives was connected to the same nonsense syllable, one member of the pair on the stimulus side, one on the response side, for example: noonday-GID, GID-midday. In List 2 this arrangement holds only for two pairs of adjectives, while the other two similar adjective pairs are connected to unlike nonsense syllables. Thus "faultless" and "perfect," "noonday" and "midday" have the same syllable for each member of the pairs, while "awkward" and "clumsy," "vacant" and "empty" have different syllables. Two dissimilar adjectives, however, are paired with a common syllable (empty-KEB, KEB-clumsy). In List 3 similar pairs were in no case connected with the same syllable. As before, different orders of presentation were arranged for each list. In these arrangements an attempt was made to avoid presentation of the same item (whether syllable or adjective, stimulus or response) twice in succession. The practice list consisted of three pairs of nonsense syllables and meaningful nouns, avoiding transfer effects which might have been caused by using adjectives in the practice list.

The conjecture being tested is that the mediated responses based on backward association in List 1 will be facilitating, those in List 3 will be inhibiting, and those in List 2 will fall between. Thus we would expect errors of recall in increasing order from List 1 to List 3.

Conditions and subjects.—Groups of 11, 12, and 13 *S*s, comparable to those used in Exp. I, served in Groups 1, 2, and 3, respectively. Essentially the same instructions as in Exp. I were used. A substitute for the memory drum was improvised from a copy holder used in typing (Remington-Rand Line-a-Time). All lists, including the practice one, were printed on a paper fastened to the copy holder. The copy holder was covered with a panel board having a narrow slit window ($\frac{1}{8}$ in. by 24 in.) through which the stimulus was presented. The exposures were made manually, the rate being controlled by a click from a 2-sec. interval timer. Exposure time was 2 sec., and 2 sec. elapsed between trials. The learning proper took place after the short practice session and was followed, as before, by backward recall from the response items typed on cards.

Results

The mean number of trials required to memorize the three pairs composing the practice list was 4.15, 4.50, and 4.55 for Groups 1, 2, and 3, respectively. The differences between these groups were not significant ($F = 0.29$, $df = 2/35$).

The mean number of trials to reach successive criteria with the lists is shown in Fig. 2. The mean number of trials to reach a criterion of one perfect trial is significantly different among lists ($F = 5.07$, $df = 2/35$, $P = .05$). In terms of t ratios, there were significant differences between

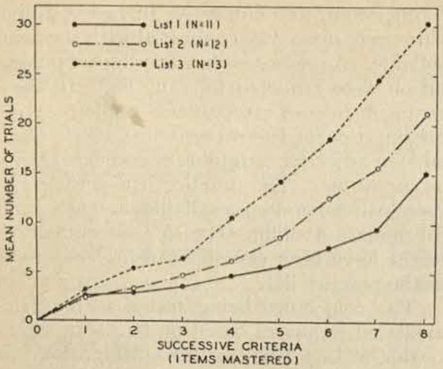


FIG. 2. Mean number of trials to successive criteria (Exp. II). (In List 1 both members of similar adjective pairs are associated with the same nonsense syllable, in List 2 half the pairs have both members associated with the same nonsense syllable, in List 3 no similar adjective pair has its members associated with the same syllable.)

Lists 1 and 3 ($t = 2.14$, $P = .05$) and between Lists 2 and 3 ($t = 2.30$, $P = .05$), though the difference between Lists 1 and 2 was not significant ($t = 1.48$).

The mean numbers of errors in the first half and second half of learning are shown in Table 6. An analysis of variance similar to that shown in Table 3 revealed only Lists as a significant effect ($F = 6.225$, $df = 2/33$, $P < .01$). The order of increasing errors by list is from List 1 with fewest errors to List 3 with most errors, as predicted.

The number of correctly reproduced responses in backward recall was 5.54 for List 1, 5.25 for List 2, and 5.07 for List 3. Thus there were no significant differences between the lists in backward recall. Even *S*s who learned List 1, which would seem quite easy to recall backward, were confused in recall, and were not protected from error by the similarities between the adjectives from the stimulus side and from the response side. All *S*s, regardless of list learned, made more backward recall errors from syllable to adjective (Means = 2.5, 2.6, and 2.1, respectively, for Lists 1, 2, and 3) than in the direction from adjective to syllable (Means = 0.0, 0.1, and 0.3, respectively).

Thus our prediction that the lists would fall in the order of difficulty from List 1 to List 3, with List 1 the easiest, is borne out, and gives further support to the interpretation that the results are mediated by backward associations.

DISCUSSION

In Exp. 1 the more common items there were on the stimulus and response sides in paired-associate learning the more difficult it was to master a list. However, early in learning this was not so, and it was in fact easier to learn the first few items of a list having more common items in S and R. We interpreted the interference as due to the formation of backward associations, and left unaccounted for the early advantage of the common-item lists.

The early advantages of the common-item lists have been found before. Umemoto (1958), in a somewhat similar experiment using group learning, found that the list with all items in common (similar to our List 1 of Exp. I) was easy early in learning, while the list with half of the items similar (corresponding to our List 2) was more difficult early in learning. This correspondence between the two studies, one done in Japan and one in the United States, implies that the result is not accidental. Hence it is

TABLE 6

MEAN CUMULATIVE ERRORS IN EARLY AND LATE STAGES OF LEARNING

List	*N*	Stage I (0–50% Learning)	Stage II[a] (50–100% Learning)
1	11	25.09	28.73
2	12	34.75	48.83
3	13	68.85	63.31

[a] Stage II vs. Stage I for pooled *S*s, $df = 1/33$, $F = 514/641 = 0.8$, *ns.*

necessary to postulate some difference in process between what is going on in the early stages of learning and what is going on in the later stages. One possibility is that before the association between S and R has progressed there is a false appearance of association because of the intrusion of some stimulus items into the responses. It has been suggested that in paired-associate learning the first stage consists in the acquisition of response items before associations between S and R develop (Morikawa, 1959a; Umemoto, 1951; Underwood, Runquist, & Schulz, 1959). If that is the case, the "familiar" item is likely to be offered as a response, regardless of the presented stimulus. But the stimulus items are also "familiar" through incidental learning, and not yet differentiated from response items. Hence some stimulus items will appear as responses. In this stage, if some of the stimulus items are the same as the response items, the independent acquisition of response items will be obscured by the intrusion of stimulus items, some of which, by chance, may get scored as "correct" acquisitions, although most of them are, in fact, interferences with learning. If there is no common item between S and R (as in List 3), evocation of the response items will not be subject to interference through the intrusion of an occasional stimulus item. This is a tentative hypothesis to explain the early advantage of List 1 over List 2, although they later cross and List 2 comes to have the advantage. Later in acquisition, the backward associations take over, and the intrusion of stimulus items becomes more generally interfering with the correct response. Thus, in retrospect, we have abandoned the second hypothesis of Exp. I in favor of the hypothesis that early in paired-associate learning some stimulus items intrude as response items not through backward association, but through failure of discrimination between stimulus and response items at this early stage.

In Exp. II, the hypothesis of mediated backward association is tested in a situation in which the backward association may facilitate learning instead of inhibiting it. This situation, however, implies two different factors, namely, the occasional facilitation and occasional inhibition by backward mediating associations. In List 3, all the backward associations, mediated by similar or identical items, are inhibiting because responses evoked by mediation are in every case different from the correct response. List 1 is more complex, in that some intruding responses are identical with the correct one while some others are only similar to the correct one. For the pairs which have adjectives in their stimulus side, such as noonday-GID in List 1, backward association will be clearly facilitating, because the response items mediated by similar adjectives from the response side (such as a backward association leading to midday-GID) are just the same nonsense syllables as the correct ones. For the pairs which have nonsense syllables on their stimulus side in List 1, such as GID-midday, the problem is rather complex, because the response adjectives mediated by identical nonsense syllables from the response side, such as the backward association GID-noonday, are only similar to the correct ones. Does a similarity between two responses facilitate the learning of a pair having the same stimulus or is it inhibiting? According to the surface of Osgood (1949) the point of the present scheme is located slightly under the zero point, hence ought to be inhibiting. According to Underwood (1949), however, the result was rather facilitating. Ishihara and Kashu (1953) also found in the A-B, A-C paradigm, in which B and C are similar, neutral, and opposite, that there were some facilitating effects when the responses are similar. From these results it is a reasonable inference that a similarity between two responses does not inhibit learning. Thus there is no inhibiting situation in List 1. Of course in List 3 of Exp. II, there are the conflicting arrangements of A-B, A-C (backward association), with no similarity between response items B and C for either nonsense syllables or adjectives, so there is no problem when we compare the *relative* difficulties of List 1 and List 3 in Exp. II.

SUMMARY

Two experiments were reported here. In Exp. 1, three lists, each consisting of eight pairs of nonsense syllables, were used. The number of common items between stimulus and response was 8 (all) for List 1, 4 (half) for List 2, and none for List 3. The more common items there were the more trials to mastery though in an early stage there was a tendency for learning a list having half common items to be more difficult than learning a list having all items of S and R in common. A hypothesis of backward mediated association was proposed.

In Exp. II, the hypothesis is tested in a situation in which the backward association facilitates learning as well as inhibiting it. The materials used were three lists all of which had the same items on the stimulus and response sides, differing only in arrangement. These lists consisted of eight pairs, half of which had adjectives as stimuli and half of which had nonsense syllables as stimuli. List 1 was so arranged that all backward mediated responses would be the same or similar to correct responses. For List 2 half of the backward associations would be facilitating, while for List 3 all backward associations were designed to be inhibiting. The results confirmed the hypothesis. The mediation through backward association thus accounts both for the inhibition in Exp. I and the facilitation and inhibition in Exp. II.

REFERENCES

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

HILGARD, E. R. Method and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

ISHIHARA, I., & KASHU, H. Learning of responses having relations of similarity, oppositeness, and neutrality. *Jap. J. Psychol.*, 1953, **24**, 1–12.

MORIKAWA, Y. Function of stimulus and response in paired-associate verbal learning. *Psychologia*, 1959, **2**, 41–56. (a)

MORIKAWA, Y. Studies in paired-associate learning: IV. On the influence of intra-stimulus and intra-response similarity upon learning and recall. *Jap. J. Psychol.*, 1959, **3**, 116–127. (b)

MURDOCK, B. B., JR. "Backward" learning in paired-associates. *J. exp. Psychol.*, 1956, **51**, 213–215.

OSGOOD, C. E. The similarity paradox in human learning: A solution. *Psychol. Rev.*, 1949, **56**, 132–144.

PRIMOFF, E. Backward and forward association as an organizing act in serial and paired-associate learning. *J. Psychol.*, 1938, **5**, 375–395.

UMEMOTO, T. On the relative weight of stimulus versus response words in rote learning. *Jap. J. Psychol.*, 1951, **21**, 46–55.

UMEMOTO, T. Similarity between stimulus and response words in paired-associate learning. *Tohoku J. exp. Psychol.*, 1958, **2**, 95–102.

UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 29–43.

UNDERWOOD, B. J. Studies of distributed practice: VIII. Learning and retention of paired-nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142. (a)

UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired-adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149. (b)

UNDERWOOD, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, **61**, 150–166.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

YOUNG, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

(Received June 27, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 105–112

JUDGMENTS OF THE RELATIVE FREQUENCY OF A SEQUENTIAL SERIES OF TWO EVENTS

DWIGHT E. ERLICK

Aerospace Medical Laboratory, Wright Air Development Division

Determining the relative frequencies of a number of mutually exclusive events that occur in a temporal sequence might require different perceptual skills as a function of (*a*) the rate of occurrence of the events, (*b*) the number of mutually exclusive events, or (*c*) the use of aids, such as counters or paper and pencil. When an individual tally cannot be made after each occurrence because of conditions such as a fast rate of presentation or the lack of tallying aids, discrimination between the relative frequencies is probably based on the sequencing of the different events with respect to themselves and each other. For example, if a series of events is presented randomly, as the relative frequency of one event is increased the probability of that event following itself in sequence will increase. Thus, a means of determining relative frequencies at fast rates of presentation might be comparison of the size and number of repetitions or clusters of the different events. Some other factors which might have a bearing on the perception of relative frequency are the location of clusters (e.g., at the beginning or end of the series), the overall length of the series, and the rate of occurrence of the events.

Hornseth and Grant (1954) found a typical psychophysical relationship between the perception of the more frequent of two sequentially presented events and the percentage difference between the events. In this study, *S*s were asked to make the judgment of relative frequency after attempting to predict the sequential series of events one at a time. The events were presented at the rate of one every 3 sec., and *S*s were not aware that they would be asked to make the overall relative frequency judgment until the entire series was completed.

Neimark and Shuford (1959) investigated the effect of estimating the overall percentage of one of two randomly occurring letters on the prediction of which of these letters would occur next in sequence. Decks of 100 cards were used with a probability of .67 for one of the letters. It was found that prediction proportions on the last 30 cards of each deck approximated .67 for the groups which only predicted, and was significantly higher for the groups which predicted *and* estimated.

Erlick (1959), in a study designed to investigate ability to estimate the percentage of occurrence of one of two random sequential events, used fast rates of presentation (5 per sec.) of the two events and instructed *S*s at the start of the series that they would be asked to identify the more frequent event and estimate its actual percentage of occurrence. It was found that when the frequencies of the events differed, *S*s, as a group, overestimated the more frequent event by approximately 2%.

The present series of studies was designed to determine how identification of the more frequent of two random sequential events is influenced by: (*a*) the relative clustering of events and the location of clusters within a series, (*b*) the length of series observed, and (*c*) the rate of occurrence of the events.

General Method

A punched paper tape reader was used to present a preprogramed series of two mutually exclusive events on a single-plane, in-line display unit. This unit utilized lenses etched with the symbols to be used. Lights behind each of the lenses projected the symbols onto approximately the same location of a 3 × 4 in. frosted plastic surface. The symbols, which were $3\frac{3}{4}$ in. high, were seen as an outline of light against a dark background. A motor driven cam supplied pulses to the tape reader which controlled the rate of presentation of the symbols; rate was varied by changing the cam. Symbols were presented one at a time and were on for approximately half of each cycle, allowing *S* to distinguish between repeated presentations of the same symbol. All of the programing equipment was housed in a separate room to reduce noise.

EXPERIMENT I: SEQUENCE

Method

The letters "N" and "S" were used to indicate two mutually exclusive events. The frequencies of occurrence of the two events were always the same for any given trial. Two different trial lengths were used: a "short" series (F) consisting of 25 "Ns" and 25 "Ss," and a "long" series (L), which contained 50 "Ns" and 50 "Ss." The sequence in which the "Ns" and "Ss" were presented to *S*s was varied in one of five different ways: 50% of the presentations of one of the symbols occurred within either the first 25%, the first 38%, 50%, 62%, or 75% of the complete sequence of "Ns" and "Ss." The short and long series and the particular sequences of the events were designated as follows: F26[1], F38, F50, F62, F76[1], L25, L38, L50, L62, and L75; the number indicates at what percentage point of the total distribution of *both* events, 50% of *one* of the events had been presented, and will be referred to as the median (Mdn.) location of that event. As an example, in the F38 condition, 13 "Ns" or 13 "Ss" occurred within the first 19 symbols displayed, while in the L62 condition, 25 of the same symbol occurred within the first 62 symbols displayed. Except for these constraints, the sequences of the two events were presented in a random order. This design has the effect of producing different relative degrees of repetition, or clustering of the events, and different locations of the clusters within the distribution. The sequences of events in the 25% and 75% conditions are exact opposites of each other, and in terms of event sequences, both produced the highest difference in clustering of the two events; i.e., in the 25% condition, the first 25% of the occurrences are repetitions of the same event, while in the 75% condition, the last 25% are repetitions of the same event. The 38% and 62% conditions are also exact opposites but there is less relative difference in clustering between the two events than for the 25% or 75% conditions, while in the F50 and L50 conditions there is no relative difference in the clustering of the events.

The sequential constraints were all counterbalanced with regard to the symbols "N" and "S." A trial consisted of viewing one of the 10 different sequential orders at the rate of 4 events per sec. This rate was well beyond the capacity of *S* to make individual tallies or counts of each symbol as it appeared. At the end of each trial, *S* was instructed to write the letter which appeared more often, and next to this letter, to put an estimate of the percentage of the time the letter appeared. The *S* was told to guess if he was not sure of the answer and to keep all previous answers covered with a card. Trials were grouped into sessions consisting of 100 trials, counterbalanced so that each of the 10 conditions preceded and followed every other condition, including itself, once. Thus, within one session, *S* made 10 observations on each of the 10 experimental conditions. All *S*s were given two sessions, separated by at least one day. The *E*, who sat in the room with *S*s, called out the number of each trial before it was presented and started the trial by pressing a button. The *S*s were given three practice trials before each session, and an attempt was made to motivate them by giving a monetary bonus as a function of "the number of correct answers." Thirty college students were paid to serve as *S*s and were run in groups of 5 to 10. Each *S* was seated in an individual booth facing the display so as to view the symbols binocularly from a distance of from 6 to 9 ft.

Results

The first session was used to acquaint *S*s with the procedure. Therefore, the following analyses were based on the second session. Two performance measures were obtained from each *S*: the number of "more frequent" responses to the letter whose median occurred at various specified locations in the distribution, and the percentage estimates of the relative frequency of occurrence of this event. To determine whether these two measures produced significantly different results, each of the measures obtained from each *S* was converted into a *T* score (McCall,

---

[1] These frequencies cannot provide 25% and 75%.

TABLE 1

P Values for Pairs of Median Locations Based on Percentage Estimates of "More Frequent" Event

Mdn. Location	Mdn. Location			
	25	38	50	62
75	.47	.11	.01	.01
62	.05	.19	.40	
50	.01	.09		
38	.06			

1939, pp. 505–508). The *T* scores of the "more frequent" responses of each *S* were then subtracted from the *T* scores of the percentage estimates. If the functional relationship of the two measures with regard to the experimental variables were the same, an analysis of variance based on these *T*-score differences would show no significant differences or interaction between Mdn. location and length of series. This hypothesis was confirmed in that no significant differences were found. Therefore, for the following analyses, only the actual percentage estimates were used, since this performance measure provided an interval value of *S*s absolute estimate of relative frequency.

Using Grant's (1956) extension of Alexander's (1946) trend analysis, an orthogonal polynomial analysis of the trends indicated a significant quadratic trend ($F = 10.24$, $df = 1/58$, $P < .005$) for the percentage estimates as a function of Mdn. location. The quartic component was also significant. There were no significant differences between F and L series, and no significant interaction between Mdn. location and the two series. Individual Wilcoxon tests comparing the five Mdn. locations with each other showed no significant differences between those conditions having the same relative degree of cluster, i.e., the 25–75% and the 38–62% Mdn. locations. However, significant differences were shown to be present between Mdn. locations having different relative degrees of clustering of the two events (see Table 1). Figure 1 indicates that for a high relative difference in the degree of clustering of two events, the event having the higher degree of cluster or repetitiveness is estimated as having the higher relative frequency of occurrence. This degree of overestimation of relative occurrence is the same when the clustering of the event is at the very beginning or the very end of the total distribution of both events.

Experiment II: Duration of Observation

Method

Observation times ranging from 2.25 to 40 sec. were presented. The observation times, the frequencies of the more frequent event (E_M) and the less frequent event (E_L), and the percentage of increment ($\Delta F = \frac{E_M - E_L}{E_L} \times 100$) for the five ranges of observation time, $T_1 - T_5$, are shown in Table 2. The observation times vary slightly within each T condition, since E_M was kept relatively constant in an attempt to obtain the same ΔFs for each T condition.

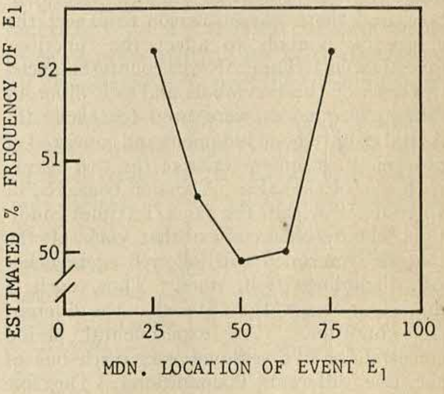


Fig. 1. Estimated frequency of one of two equally frequent events as a function of the median location of one of the events within the distribution.

TABLE 2

Observation Times, Frequencies of Events, and Percentage Increments of More Frequent Event

Observation Time (Sec.)		Number of Events		$\Delta F = \frac{E_M - E_L}{E_L} \times 100$
		E_M	E_L	
T_1	2.50	5	5	0
	2.75	6	5	20
	2.25	5	4	25
	2.50	6	4	50
T_2	5.00	10	10	0
	4.75	10	9	11.1
	4.50	10	8	25.0
	4.25	10	7	42.9
T_3	10.00	20	20	0
	9.50	20	18	11.1
	9.00	20	16	25.0
	8.50	20	14	42.9
T_4	20.00	40	40	0
	19.00	40	36	11.1
	18.00	40	32	25.0
	17.00	40	28	42.9
T_5	40.00	80	80	0
	38.00	80	72	11.1
	36.00	80	64	25.0
	34.00	80	56	42.9

The symbols "1" and "2" were used to indicate the two mutually exclusive events. (These symbols differed from those in Exp. I and Exp. III because other experiments being carried on simultaneously with this equipment made it inconvenient to use the same symbols, and there was no reason to expect the choice of symbols to affect the functions investigated.) Each ΔF was counterbalanced in terms of these symbols and two different random sequences were used for each ΔF. A trial required one judgment and consisted of observing a random series of the two events with one of the ΔFs. A session consisted of 16 trials all within the same T (time) condition, counterbalanced so that each of the four ΔFs preceded and followed every other ΔF, including itself, once. Thus, each *S* made four observations at each of the different ΔF conditions. An experimental period consisted of five sessions, each with one of the five different T conditions. The five conditions were presented to five groups of 20 *S*s each, according to a latin square design. In order to determine if a few observations averaged over many *S*s produced the same functional relationship as many observations made by one *S*, two *S*s from the group experiment were also run independently of the foregoing and made 100 observations at each of the experimental conditions according to a similar latin square design. The selection of these two *S*s was based on their availability and their apparent high motivation, as reflected by attentiveness during the group testing period and their consistency of performance, i.e., a systematic increase in accuracy as ΔF increased. Because there was no significant difference between the number of more frequent responses and percentage estimates in Exp. I, it was possible to simplify data reduction by using only the more frequent responses in Exp. II and III. Similar instructions and the same rate of presentation of events were used as in Exp. I.

Results

The ΔF required for 75% correct identification of E_M furnished a measure of the relative effect of the different lengths of observation. This measure was obtained by linear interpolation and for the group was obtained by averaging the responses of all *S*s over each of the observation times. The ΔF = 0 condition gives a measure of symbol bias. There was no systematic interaction between symbol preference and duration of observation. It is apparent that the results of the two independent *S*s and of the group were similar (see Table 3 and Fig. 2); first there was a decrease, and then an increase in accuracy of performance, as the duration of observation was increased from 2.25 to 40 sec. The overall increase in performance is equivalent to a change in ΔF of approximately 10 (e.g., a drop from 20 ΔF to 10 ΔF for S_2). An analysis of variance of the group data based only on correct responses at ΔF = 25 showed the main effects of duration to be significant ($F = 5.44$, $df = 4/380$, $P < .001$). The initial decrease in accuracy between T_1 and T_2 was also found to be significant ($P < .01$), based on Tukey's test for multiple comparisons (Ryan, 1959).

The overall effect of length of observation on the percentage of correct responses depends to some degree on the ΔF increment. At $\Delta F = 11.1$, there is no apparent effect for the grouped data, while at $\Delta F = 25.0$, the group and the independent *S*s show an initial decrease and then an increase in accuracy of responses; at $\Delta F = 42.9$, one independent *S* still shows an initial decrease and then an increase in accuracy, while the other independent *S* and the group appear to have reached an asymptote in terms of correct identifications of the more frequent event.

EXPERIMENT III: RATE OF PRESENTATION

Method

The conditions T_2 and T_4 (see Table 3) were presented at four different rates: 1, 2, 4, and 8 events per sec. Within each of the two T conditions, four groups, ranging from 7 to 11 *S*s each, were presented the different rates according to a latin square design. The T_2 and T_4 conditions had totals of 35 and 37 *S*s, respectively. The letters "A" and "B" were used to designate the two mutually exclusive events. As in Exp. II, the symbols and the sessions were counterbalanced and *S*s were required to indicate only the "more frequently" occurring symbol. Instructions similar to those in Exp. I and II were used.

TABLE 3

PERCENTAGE CORRECT JUDGMENTS AS A FUNCTION OF OBSERVATION TIME

Duration	Group (G) or *S*s	ΔF			
		0[a]	11.1	25.0	42.9
T_1	G	53.8	63.2[b]	83.8	84.1[b]
	S_1	58.0	72.4	88.0	90.1
	S_2	63.0	70.2	87.0	90.6
T_2	G	50.0	67.5	74.5	85.5
	S_1	45.0	61.0	90.0	93.0
	S_2	54.0	62.0	81.0	97.0
T_3	G	47.2	65.2	79.0	84.2
	S_1	47.8	66.0	82.0	86.0
	S_2	57.8	66.0	89.0	98.0
T_4	G	48.8	66.3	83.2	82.8
	S_1	42.0	69.0	88.0	91.0
	S_2	58.0	80.0	88.0	95.0
T_5	G	50.5	68.7	86.0	84.2
	S_1	56.0	78.0	94.0	93.0
	S_2	58.5	87.0	94.0	98.0

[a] Percentage of number "1" responses.

[b] At T_1 the ΔF values for 11.1 and 42.9 are interpolated.

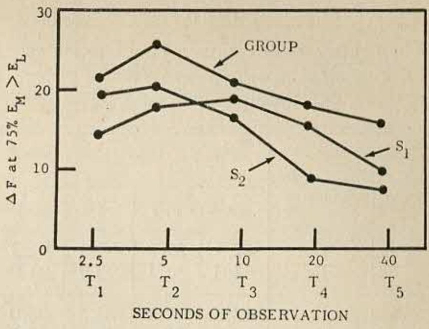


FIG. 2. Effect of duration of observation on ΔF at 75% correct identification. (The abscissa is scaled logarithmically.)

Results

As in Exp. II, the ΔF required for the interpolated 75% correct identification of E_M was used as a measure of performance. Results shown in Table 4 and Fig. 3 indicate an overall increase in the ΔF required for 75% accuracy as a function of increasing rate of presentation for both durations of observation T_2 and T_4. An analysis of variance based on correct responses at $\Delta F = 11.1$ showed the main effect of rate over both T_2 and T_4 to be significant ($F = 15.35$, $df = 3/180$, $P < .001$) and the interaction between the T conditions and rate significant ($F = 5.73$, $df = 3/180$, $P < .001$). This indicates that the effects of rate are different for the short (T_2) and long (T_4) series of events. At the slowest rate of 1 per sec., T_2 requires a smaller ΔF than that of condition T_4 to obtain 75% accuracy; while

TABLE 4

PERCENTAGE OF CORRECT JUDGMENTS AS A FUNCTION OF OBSERVATION TIME AND PRESENTATION RATE

Rate (Sec.)	Observation Time	ΔF			
		0[a]	11.1	25.0	42.9
1	T_2	47.9	87.9	89.3	94.3
	T_4	54.0	81.1	93.9	91.9
2	T_2	43.6	77.9	83.6	95.0
	T_4	56.1	77.0	86.5	87.8
4	T_2	49.3	70.0	82.9	90.7
	T_4	46.0	73.0	87.2	88.5
8	T_2	56.4	52.4	88.6	92.1
	T_4	71.6	70.3	85.8	87.8

[a] Letter "A" responses.

as the rate of presentation is increased, the ΔF required for T_2 to achieve the same accuracy as T_4 becomes increasingly greater.

Here again, as in Exp. II, the results are influenced by the degree of ΔF increment. At ΔF = 11.1, there is a significant interaction between the length of observation and rate of presentation. However, as ΔF is increased to 25 and 42.9, this interaction disappears, *S*s apparently reach an approximately equivalent asymptote for both durations of observation.

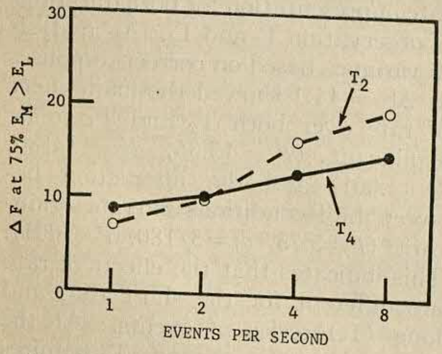


FIG. 3. Effect of rate of presentation on ΔF at 75% correct identification. (The abscissa is scaled logarithmically.)

DISCUSSION

Sequence.—The results of Exp. I substantiate the hypothesis that the sequence in which a given series of two events is viewed influences the a posteriori judgment of the relative frequency of occurrence of the events. Thus, when the frequency of occurrence of two events is actually the same, the event having the higher degree of clustering is estimated as having a higher relative frequency of occurrence. Also, the fact that there was not a significant difference between the lengths of the two different series indicates that the important variable appears to be the *relative* difference in the degree of repetition of the two events, and not the *absolute* degree of repetition of an event. This effect appears to be the same whether the repetition of the event occurs at the very beginning or at the end of the distribution of both events. Whether these results would be extended to the condition where repetition of an event occurs elsewhere in the distribution requires further investigation. Because of the influence of highly repetitive sequences of the same stimulus on the perception of the more frequent event found in Exp. I, more than one randomization was used for the same condition in Exp. II and III.

Duration of observation.—When an observer attempts to discriminate between the relative frequencies of events that are presented too fast to be tallied separately, he probably utilizes clusters, i.e., reorganizes the input sequences into larger units or "chunks," which are used as the basic unit for tallying (Miller, 1956). Thus, the relative length and number of "chunks" of the different events are probably what is being differentially weighted. This might explain the nonmonotonic function relating accuracy of perceiving the more frequent event to duration of observation. The very shortest duration approximates immediate recall, in terms of the number and size of "chunks." Thus, the task becomes essentially one of discrimination. However, as the duration of observation increases, the task of recalling previous details of the series becomes greater. At the same time, however, an increase in observation time gives *S* a larger sample upon which to base his judgment. It may be the interaction between the foregoing that causes the nonmonotonic relationship in Exp. II.

The close similarity between the individual *S* curves (based on 100 observations at each point on the curve) and the group curve (based on observations by each of 100 different *S*s) adds validity to the general functional relationship obtained in this study. Because of this similarity and the relative ease of obtaining group data, the study involving the effects of rate was conducted using only groups of *S*s.

Rate of presentation.—Increasing the rate of presentation probably has the effect of increasing the number of events that are grouped together in "chunks" as a tallying unit. Since *S* can make only a finite number of discriminations per unit time, if more information is forced into a specified time period, the information is probably broken down into larger "chunks" of the sequence with a consequent loss. Therefore, given the same amount of information, as its rate of presentation is increased, the ability to assimilate the information will tend to decrease. The foregoing might explain the general decline in accuracy of perceiving the more frequent event as rate of presentation was increased from 1 to 8 events per sec.

The higher accuracy for the shorter series at the slowest rate might be explained in terms of the interaction between memory and the increased time sample. Thus, at the rate of 1 event per sec., individual events can probably be discriminated one at a time, and the premium for accuracy would tend to favor memory functions resulting in greater accuracy for a shorter series. On the other hand, at fast rates of 4 and 8 events per sec., where some information is lost because of the necessity for grouping the individual events, increasing the length of the series, with a consequent increase in time sample, appears to outweigh the advantage of memory that accrues to the shorter series.

In general, it appears that approximately a 10 to 20% increment (ΔF) of the more frequent event is required to maintain a 75% accurate identification of that event, when the observation time is greater than 5 sec. and the rate of presentation is not faster than 8 events per sec.

The problem of perceiving the more frequent or modal event of a sequential series of two events has been shown to be a function of the sequence in which the events appear, the total number of events observed, and the rate at which the events are presented with rate interacting with observation time.

SUMMARY

The effects of sequence, duration of observation, and rate of occurrence on the perception of the more frequently occurring event of a sequential series of two mutually exclusive events was investigated. It was found that: when both events are equally frequent, the event having the higher degree of clustering is estimated as having a higher relative frequency of occurrence; the relative degree of cluster—not the absolute magnitude of the cluster—appears to influence the perception of relative frequency; as duration of observation increases from very short time periods, the 75% threshold of identifying the more frequent event decreases and then increases; as rate of presentation increases, this threshold decreases; rate interacts as a function of duration of observation, i.e., the shorter the duration, the steeper the decline in accuracy as a function of increasing rate; and an increment (ΔF) of approximately 10 to 20% of the more frequent event (E_M) is required to maintain 75% accurate identification of E_M with observation times beyond 5 sec. and rates of presentation 8 per sec. or slower. The results are explained in terms of the discrimination of varying size clusters of events or "chunks" of the sequence of events.

REFERENCES

ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.

ERLICK, D. E. Judgments of the relative frequency of sequential binary events:

Effects of frequency differences. *USAF WADC tech. Rep.*, 1959, No. 59–580.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Hornseth, J. P., & Grant, D. A. The discrimination of random series of stimulus frequencies as a function of their relative and absolute values. *USAF AFPTRC tech. Rep.*, 1954, No. 54–76.

McCall, W. A. *Measurement.* New York: Macmillan, 1939.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Neimark, E. D., & Shuford, E. H. Comparison of predictions and estimates in a probability learning situation. *J. exp. Psychol.*, 1959, **57**, 294–298.

Ryan, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56**, 26–47.

(Early publication received January 3, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 113–121

MEDIATING VERBAL RESPONSES AND STIMULUS SIMILARITY AS FACTORS IN CONCEPTUAL NAMING BY SCHOOL AGE CHILDREN[1]

HARVEY M. LACEY[2]

University of Massachusetts

It has been shown in a number of recent studies (e.g., Baum, 1951; Cary & Goss, 1957; Fenn & Goss, 1957; Goss & Moylan, 1958; Lacey & Goss, 1959) that prior learning of potential verbal mediating responses facilitates subsequent conceptual behavior. In this experiment acquisition of conceptual naming responses was investigated as a function of patterns of relationships among initiating stimuli, verbal mediating responses and stimuli, and verbal terminating responses (conceptual naming), and of similarity among the initiating stimuli. Four patterns of possible relationships were included; and the design provided for determination of their separate effects as well as for comparisons of those effects.

Figure 1 shows the four patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses. Pattern AA has been labeled acquired equivalence of cues (Dollard & Miller, 1950; Goss, 1955). In this pattern, acquisition of a common verbal response to two or more initiating stimuli is followed by acquisition of a common terminating response.

Acquired distinctiveness of cues is the label commonly applied to Pattern BB (Dollard & Miller, 1950; Goss, 1955). Dissimilar verbal mediating responses are learned to each of two or more initiating stimuli. Subsequently, discriminative terminating responses are conditioned to the initiating stimuli to which the discriminative mediating responses had been learned.

Although the existence of Patterns AB and BA has been suggested previously (Goss, 1955), neither pattern has been included in prior investigations. In Pattern AB a common mediating response is learned to two or more initiating stimuli, after which different terminating responses are learned to those stimuli. In Pattern BA, different mediating responses are

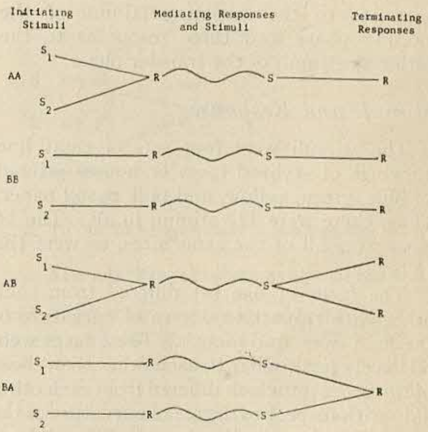


FIG. 1. Four patterns of relationships among initiating stimuli, common or discriminative mediating responses and stimuli, and common or discriminative terminating responses.

---

[1] This study is based on a dissertation presented in partial fulfillment of the requirements for the PhD at the University of Massachusetts. The writer would like to express his appreciation to Albert E. Goss for his critical suggestions and helpful guidance, and to the other members of his committee, Solis Kates, Claude C. Neet, and Raymond Wyman, for their counsel.

[2] Now at the Pittsburgh Child Guidance Center, Pittsburgh, Pennsylvania.

conditioned to initiating stimuli, and a common terminating response is then conditioned to the initiating stimuli.

METHOD

Design

Control conditions.—The control for warm-up and learning-performance sets was by means of a condition in which common or discriminative responses were first conditioned to sets of initiating stimuli which did not resemble the stimuli of the transfer phase. However, features of the training and transfer phases other than the sets of initiating stimuli employed were the same. Therefore, those experiences involved in warm-up and learning-performance sets were expected to generalize from the training phase to the transfer phase as was facilitation based on receptor-orienting responses.

Combinations of conditions.—Under the experimental condition each of the AA, BB, AB, and BA patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses was combined both with the more similar and with the less similar initiating stimuli. Corresponding to each of these eight combinations was a control for the combined effects of nonspecific transfer and receptor-orienting responses. Thus, there were 16 combinations of conditions (Fig. 2). Pattern AA was combined with Pattern BB as was Pattern AB with Pattern BA so that all *S*s learned three responses to the initiating stimuli of the training phase and three responses to the initiating stimuli of the transfer phase.

Stimuli and Responses

The stimuli were four sets of eight line drawings of stylized faces or houses printed on blue, green, yellow, and pink pastel paper. Thus, there were 128 stimuli in all. The 16 faces were all of the same sizes, as were the 16 houses.

The faces of one set differed from each other with respect to degree of curvature of the hair, eyes, and mouth. These faces were relatively less similar to each other than those of the other set which differed from each other only with respect to degree of curvature of the eyes. The less similar houses differed from each other with respect to direction of smoke from the chimney and in number of scalloped bottoms of the shades in the two windows and of the shade in the window of the door. The more similar set of houses differed only with respect to number of the scalloped bottoms of the shades in the two windows.

The six nonsense syllables used were selected from among those of Mandler's (1955) list which have high associative frequencies.

Apparatus

The apparatus consisted of a vertical panel in which a window (5 in. × 3 in.) had been cut. Behind the window, guides correctly positioned the 3 × 6 in. plastic cards on which the stimuli appeared. When a card was dropped between the guides, the bottom of the card tripped a microswitch which activated a time-delay relay set for 3 sec. At the end of the anticipation interval a bulb on the front of the apparatus lighted.

Subjects

The *S*s were 96 children who were drawn from among those between the ages of 8 and 11 yr. in Grades 2 through 6.[3] The groups were equated for sex and as nearly as possible for means and ranges of ages and for IQs. All children with IQs of less than 90 or more than 125 were excluded.

Procedure

The same sequence of training and transfer experiences was administered to each of the eight groups. During the training phase, *S*s were familiarized with the mediating responses before those responses were conditioned to the initiating stimuli of Patterns AA and BB or of Patterns AB and BA. The *S*s were then familiarized with the terminating responses of the transfer task after which those responses were associated with the initiating stimuli.

In order to counterbalance for any differences due to experience with face stimuli vs. house stimuli, half of the *S*s under the experimental condition had faces for both transfer and training phases; the other half of each group had houses for both training and transfer phases. Under the control condition, half of the *S*s had faces for the training phase and houses for the transfer phase; the other half of each group had houses for the training phase and faces for the transfer phase.

Familiarization with mediating responses.—Familiarization training began by *E* telling each *S* that he was to say some short syllables

---

[3] The cooperation and kind permission of the Erving School Union, William Trainor, Superintendent of Schools, is gratefully acknowledged.

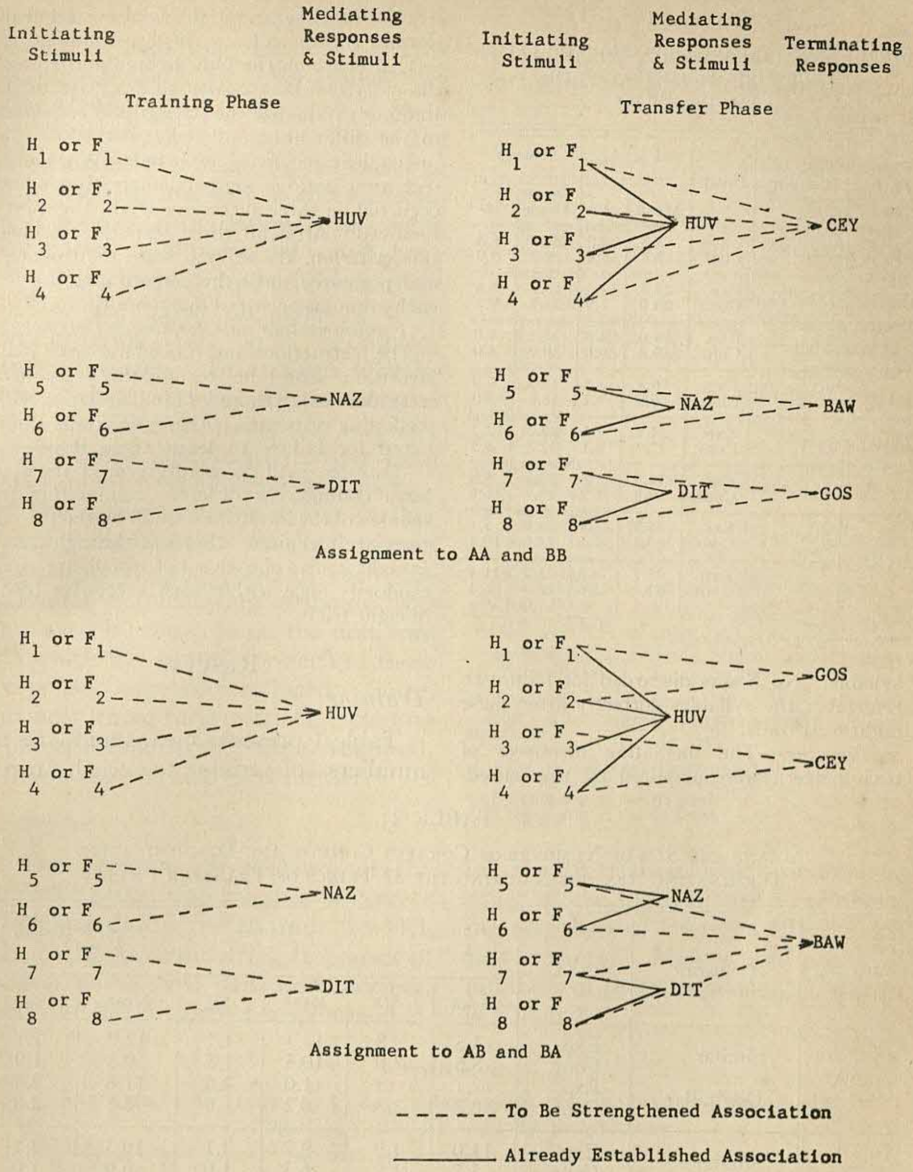


FIG. 2. Relationships among initiating stimuli, mediating responses and stimuli, and terminating responses for assignment to Patterns AA and BB or to Patterns AB and BA with dissimilar initiating stimuli. (The relationships with similar stimuli are the same. Under the experimental condition, the initiating stimuli of both the training phase were identical. Under the control condition, the face or house initiating stimuli of the training phase are replaced by house or face stimuli, respectively, for the training phase.)

in the same way that *E* said the syllables. The *S* was also told to try to memorize them. Each of the three syllables was then presented three times, once within three successive three-trial blocks. At the end of nine trials *E* asked *S* to name the syllables. Those who were unable to remember them were given six additional trials, two with each

TABLE 1

Means and Standard Deviations of Trials to the First Criterion of the Training Phase

Pattern	Similarity	Cond.	Faces		Houses	
			Mean	*SD*	Mean	*SD*
AA	Similar	Exp.	15.8	8.9	21.8	6.3
		Cont.	18.8	5.9	19.0	8.0
	Dissimilar	Exp.	22.8	12.4	18.7	5.6
		Cont.	20.0	7.9	22.7	9.8
AB	Similar	Exp.	19.0	3.8	21.3	8.0
		Cont.	19.8	9.0	20.2	8.0
	Dissimilar	Exp.	16.8	3.5	18.3	6.9
		Cont.	16.8	9.1	13.3	5.0
BB	Similar	Exp.	24.2	7.1	27.3	12.2
		Cont.	27.7	8.1	32.3	5.5
	Dissimilar	Exp.	32.3	11.5	28.5	5.8
		Cont.	19.8	7.2	28.7	11.1
BA	Similar	Exp.	25.8	7.1	31.0	5.1
		Cont.	29.8	7.5	24.0	10.2
	Dissimilar	Exp.	25.2	4.5	28.7	11.4
		Cont.	27.2	5.6	25.3	6.4

Note.—*N* = 6 for all cells.

syllable. No *S* was discarded for failure to remember the syllables correctly after these additional trials.

Training.—The mediating responses of each pattern were acquired by the paired-associate technique. Each trial consisted of the presentation of one of the stimuli for 3 sec., after which the bulb lighted and *E* said the syllable which *S* was to associate to it. In order to equalize the strengths of responses to the different stimuli, the method of adjusted learning was employed. When correct anticipations for particular stimuli had occurred on three successive trials, they were removed. After all three responses reached this criterion, the stimuli were reintroduced and presented until the correct response for each stimulus occurred one more time.

Familiarization with terminating responses.—The instructions and procedures for familiarization with the terminating responses were identical to those for familiarization with mediating responses. Again, no *S* was eliminated for failure to learn these responses.

Transfer.—All *S*s were given 32 trials to learn common or different terminating responses to the initiating stimuli by the paired-associate technique. Each of the eight faces or eight houses of each set of stimuli appeared randomly once within each successive block of eight trials.

RESULTS

Training

Table 1 presents means and *SD*s of numbers of trials to reach three

TABLE 2

Means and *SD*s of Numbers of Correct Common and Discriminative Terminating Responses during the 32 Trials of Transfer Phase

Pattern	Similarity	Cond.	Faces *N* = 6 ea.		Houses *N* = 6 ea.		Faces & Houses *N* = 12 ea.	
			Mean	*SD*	Mean	*SD*	Mean	*SD*
AA	Similar	Exp.	13.5	.8	12.3	1.1	12.9	1.1
		Cont.	8.5	2.0	10.5	1.3	9.5	1.9
	Dissimilar	Exp.	12.7	.3	11.0	2.6	11.8	2.4
		Cont.	8.8	2.3	8.2	1.6	8.5	2.0
BB	Similar	Exp.	11.0	1.9	9.7	2.1	10.3	2.1
		Cont.	5.7	1.0	6.3	1.0	6.0	1.0
	Dissimilar	Exp.	12.3	1.8	11.3	1.1	11.8	1.6
		Cont.	10.0	1.2	9.0	2.1	9.5	1.8
AB	Similar	Exp.	5.7	1.2	5.8	2.0	5.8	1.7
		Cont.	8.0	2.3	7.0	1.2	7.5	1.9
	Dissimilar	Exp.	8.0	1.0	7.8	2.0	7.9	1.6
		Cont.	9.0	2.0	8.2	2.0	8.6	2.0
BA	Similar	Exp.	7.7	.8	8.3	.8	8.0	.8
		Cont.	8.0	2.6	9.3	1.4	8.7	2.2
	Dissimilar	Exp.	6.0	1.9	6.2	.4	6.1	1.4
		Cont.	9.2	1.9	8.5	1.6	8.8	1.8

successive correct anticipations of the common mediating response for Patterns AA and AB. Also shown are the comparable means and *SD*s of numbers of trials to learn each of the discriminative mediating responses of Patterns BB and BA to three successive correct anticipations. Since the task was not mastered until both responses had reached criterion, these means are based on sums of trials to criterion for both discriminative responses. An analysis of variance among these means showed that the common mediating response was learned significantly faster than discriminative mediating responses.

The second criterion of one further correct anticipation of the common mediating response and of each discriminative mediating response was achieved by most *S*s on the first trial on which the stimuli for each of these responses were reintroduced. Since, at most, three further trials, including the criterion trial, were required, these were not analyzed further.

TABLE 3

Analysis of Variance of Correct Responses during the 32 Trials of Transfer Phase

Source	*df*	*MS*	*F*
Between *Ss*[a]	95		
Faces-Houses (B)	1	3.79	1.03
Similar-Dissimilar (C)	1	14.63	3.99*
AABB-ABBA (D)	1	273.13	74.42**
Experimental-Control (E)	1	43.13	11.75**
Error (b)	80	3.67	
Between *Ss*[b]	47		
Faces-Houses (B′)	1	6.51	2.05
Similar-Dissimilar (C′)	1	12.76	4.01
Experimental-Control (E′)	1	270.01	84.91**
B′E′	1	14.26	4.48*
Error (b)	40	3.18	—
Within *Ss*	48		
AA-BB (A′)	1	38.76	10.59**
A′C′	1	75.26	20.56**
Error (w)	40	3.66	—
Total	95		
Between *Ss*[c]	47		
Faces-Houses (B″)	1	.04	.01
Similar-Dissimilar (C″)	1	3.38	.82
Experimental-Control (E″)	1	51.04	12.24**
Error (b)	40	4.17	—
Within *Ss*	48		
AB-BA (A″)	1	5.04	1.68
A″C″	1	37.50	12.50**
A″C″E″	1	15.04	5.01*
Error (w)	40	3.00	—

Note.—The omitted interactions were all nonsignificant.

[a] For all *S*s.

[b] For *S*s assigned to Patterns AA and BB.

[c] For *S*s assigned to Patterns AB and BA.

* Significant at the .05 level.

** Significant at the .01 level.

Transfer

The performance measure of the transfer phase was number of correct responses during the 16 trials in which the common terminating responses of Patterns AA and BA were correct, and number of correct responses during the 16 trials in which one or the other of the discriminative terminating responses of Patterns BB and AB was correct. Means and *SD*s of numbers of these responses are shown in Table 2.

TABLE 4

Analyses of Variance of Correct Responses during Transfer Phase for Patterns AA, BB, AB, and BA Separately

Source	*df*	Pattern AA		Pattern BB		Pattern AB		Pattern BA	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Experimental-Control (A)	1	136.69	33.60**	133.34	25.16**	16.52	4.63*	35.02	10.74**
Similar-Dissimilar (B)	1	13.02	3.21	75.00	14.15**	30.69	8.60**	9.19	2.82
A × B	1	.02		12.00	2.26	4.52	1.27	13.02	3.99
Error	44	4.06		5.30		3.57		3.26	

* Significant at the .05 level.

** Significant at the .01 level.

Because Patterns AA and BB were administered to half of the *S*s, and Patterns AB and BA were administered to the other half, the analysis of variance to determine the significance effects of the five factors was somewhat unusual. The "between *S*s" sources of variance, in which all *S*s were involved, were: (*a*) faces or houses, (*b*) more similar or less similar stimuli, (*c*) experimental or control conditions, and (*d*) assignment to Patterns AA and BB or to Patterns AB and BA.

For *S*s assigned to Patterns AA and BB, the "within *S*s" sources of variance were Pattern AA or Pattern BB and the interactions of this factor with faces or houses, more similar or less similar stimuli, and experimental or control conditions. Each of the latter three factors and their interactions with each other were "between *S*s" sources, but now for only half of the *S*s. The same was true for *S*s assigned to Patterns AB and BA.

Patterns AA, BB, AB, and BA separately.—Nonsignificant *F*s indicated that the sets of face and house stimuli were essentially equivalent (Table 3). Accordingly, this factor was not included in analyses of variance for Patterns AA, BB, AB, and BA separately, for each of which the sources of variation were more similar or less similar stimuli, experimental or control conditions and the interaction of these two variables.

For Pattern AA, significantly more correct responses were made under the experimental condition than under the control condition. While more correct responses were made with similar initiating stimuli than with dissimilar initiating stimuli the *F* for this difference was not significant.

For Pattern BB, there were significantly more correct discriminative terminating responses under the experimental condition (Mean = 11.08) than under the control condition (Mean = 7.75), and with less similar initiating stimuli (Mean = 10.67) than with more similar initiating stimuli (Mean = 8.17).

For Pattern AB, significantly more correct discriminative terminating responses occurred under the control condition (Mean = 8.04) than under the experimental condition (Mean = 6.88), and with less similar stimuli (Mean = 8.25) than with more similar stimuli (Mean = 6.62).

For Pattern BA, the mean of 8.75 for the control condition was significantly greater than the mean of 7.04 for the experimental condition. While more correct responses occurred with more similar stimuli (Mean = 8.34) than with less similar stimuli (Mean = 7.46), the difference was not significant.

TABLE 5

ANALYSES OF VARIANCE OF CORRECT RESPONSES DURING TRANSFER PHASE FOR PATTERNS AA AND BA AND FOR PATTERNS BB AND AB

	df	Patterns AA and BA		Patterns BB and AB	
		MS	*F*	*MS*	*F*
Experimental-Control (A)	1	16.66	4.55*	27.09	6.10*
Similar-Dissimilar (B)	1	22.04	6.02*	102.09	22.99**
Patterns AA and BA (C) or Patterns BB and AB (C′)	1	187.04	51.10**	94.01	21.17**
A × B	1	7.05	1.93	1.27	.29
A × C or A × C′	1	155.05	42.36**	123.77	27.88**
B × C or B × C′	1	.17	.05	4.60	1.04
A × B × C or A × B × C′	1	5.99	1.64	14.24	3.21
Error	88	3.66		4.44	

* Significant at the .05 level.
** Significant at the .01 level.

Pattern AA and Pattern BA.—In the analysis of variance for Pattern AA and Pattern BA (Table 5), significantly more correct responses occurred under the experimental condition (Mean = 9.71) than the control condition (Mean = 8.88), with more similar (Mean = 9.78) than with less similar stimuli (Mean = 8.81), and for Pattern AA (Mean = 10.69) than for Pattern BA (Mean = 7.90).

Pattern BB and Pattern AB.—For Patterns BB and AB (Table 5), significantly more correct responses occurred under the experimental condition (Mean = 8.98) than under the control condition (Mean = 7.92), with less similar stimuli (Mean = 9.46) than with more similar stimuli (Mean = 7.40), and for Pattern BB (Mean = 9.42) than for Pattern AB (Mean = 7.46).

DISCUSSION

In agreement with Lacey and Goss' (1959) finding of an inverse relationship between acquisition rate and number of responses, the common mediating response of the training phase was acquired more rapidly than discriminative mediating responses. Of greater consequence was the finding that the parallel sets of more similar faces and of more similar houses did not differ in difficulty; this was also true for the parallel sets of less similar faces and of less similar houses.

Patterns AA, BB, AB, and BA separately.—For Pattern AA, in which acquisition of a common mediating response was followed by acquisition of a common terminating response, analyses of the roles of cue-producing responses in learning (Goss, 1955; Lacey & Goss, 1959) lead to predictions of more correct responses under experimental than under control conditions and with more similar than with less similar stimuli. Both of the obtained differences were in the predicted direction, but only the former was significant. Since the interaction of these two variables was not significant, the effects of a common mediating stimulus on acquisition of a common terminating response were apparently no different with less similar than with more similar stimuli. However, other factors may have precluded a significant interaction. For example, the mediating stimuli may have been such large proportions of compounds of initiating and mediating stimuli that even relatively large differences in the similarity of initiating stimuli could produce only relatively small differences in similarity of compounds. Or, the differences in relative similarity of the more similar to the less similar stimuli may not have been sufficient to produce a significant interaction. These same factors might also account for the nonsignificant interactions of experimental or control conditions with more similar or less similar stimuli for Patterns BB, AA, and BA separately.

For Pattern BB, discriminative mediating responses were acquired to either more similar or less similar stimuli; subsequently discriminative terminating responses were acquired. Predictions of more correct discriminative terminating responses under the experimental condition than under the control condition and with less similar than with more similar stimuli were both confirmed.

For Pattern AB, a common mediating response was conditioned to the initiating stimuli and then discriminative terminating responses were conditioned to the same stimuli. In agreement with predictions, more correct terminating responses occurred under the control than under the experimental condition and with less similar than with more similar stimuli.

Acquisition of the different mediating responses of Pattern BA was followed by acquisition of a common terminating response. As expected, more rapid learning was obtained under the control than under the experimental condition and with more similar than with less similar stimuli. However, the latter difference was nonsignificant.

Three further features of these findings are of interest. First, whatever the reason, the presumed presence of mediating stimuli under the experimental condition did not produce significant interactions with similarity of initiating stimuli. Second, all four patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses influenced conceptual behaviors. In any concept-formation situation in which two or more of these patterns are present, therefore, resultant performance probably reflects their joint influence. Third, the presence of verbal mediating responses per se does not assure facilitation of learning and performance of subsequent conceptual naming: whether facilitation or inhibition occurs is contingent on the patterns or combinations of patterns. Such facilitation or inhibition would be added to or subtracted from expected facilitative effects of warm-up and learning-performance set and of receptor-orienting responses.

Pattern AA with Pattern BB; *Pattern AB with Pattern BA*.—As predicted on the basis of previous findings (Lacey & Goss, 1959) of an inverse relationship between acquisition rate and the number of responses, the common terminating response of Pattern AA was mastered more rapidly than the discriminative terminating responses of Pattern BB. Since a greater number of correct responses occurred with the more similar than with the less similar initiating stimuli for Pattern AA, and the converse for Pattern BB, the expected interaction of Pattern AA or Pattern BB with similarity of initiating stimuli was obtained.

The common terminating response of Pattern BA was, as predicted, correct more often than the discriminative terminating responses of Pattern AB, but the difference was not significant. However, the prediction of more correct responses with less similar than with more similar stimuli for Pattern AB and the converse for Pattern BA was confirmed.

Pattern AA with Pattern BA; *Pattern BB with Pattern AB*.—The facilitative and inhibitory effects for Pattern AA and Pattern BA, respectively, produced the expected interaction of these patterns with experimental or control conditions. Facilitative effects for Pattern AA outweighed inhibitory effects for Pattern BA. Similarity of initiating stimuli did not have significant effects with Patterns AA or BA separately. For the two patterns combined, however, significantly more correct responses occurred with more similar than with less similar stimuli.

Facilitative effects with Pattern BB and inhibitory effects with Pattern AB produced the expected interaction of these patterns with experimental or control conditions with facilitation exceeding inhibition.

Comparisons involving Patterns AA, BB, AB, and BA.—Relative degree of similarity of initiating stimuli produced differences in the predicted direction with all four patterns. Specifically, more correct terminating responses occurred with more similar stimuli than with less similar stimuli for Patterns AA and BA and with less similar stimuli than with more similar stimuli for Patterns BB and AB. Only for Patterns BB and AB, however, were the differences between more similar and less similar stimuli significant. Since both patterns involved discriminative terminating responses, similarity of initiating stimuli may have relatively greater effects on transfer tasks involving discrimination than on those involving generalization.

For all 16 combinations, the largest number of correct terminating responses was obtained for Pattern AA with similar initiating stimuli under the experimental condition. The smallest number of correct terminating responses occurred for Pattern AB with similar initiating stimuli under the experimental condition.

In general, therefore, the observed relationships support inferences from previous findings that verbal mediating responses and stimuli play an important role in conceptual naming. Also, inferences of the importance of verbal mediating responses can be extended to initiating stimuli of varying relative similarity which are more similar to pictorial representations of common objects, to four separate patterns of relationships involving mediating responses, and to children.

SUMMARY

The significance of verbal mediating responses and stimuli in conceptual naming was investigated by means of a design in which four patterns of possible relationships among initiating stimuli, mediating stimuli and responses, and terminating responses were combined with sets of more similar or less similar initiating stimuli. The patterns involved initiating stimuli to which *S*s learned: (*a*) a common mediating response and then a common terminating response (Pattern AA), (*b*) discriminative mediating responses and then discriminative terminating responses (Pattern BB), (*c*) a common mediating response and then discriminative terminating responses (Pattern AB), and (*d*) discriminative mediating responses and then a common terminating response (Pattern BA).

The sets of more similar initiating stimuli consisted of stylized line-drawings of faces or of houses which differed along but one dimension. The sets of less similar faces or houses differed along three dimensions. Ninety-six children, ages 8 to 11, in Grades 2–6, were assigned to eight groups of 12 *S*s each, equated as nearly as possible for age, sex, and IQ.

Comparisons of experimental and control conditions indicated that the predicted positive transfer had occurred with both Pattern AA and Pattern BB. Faster learning was obtained with more similar stimuli for Pattern AA and with dissimilar stimuli for Pattern BB. Also confirmed were predictions of negative transfer under the experimental condition with Pattern AB and Pattern BA and of faster learning with more similar stimuli for Pattern BA and with less similar stimuli for Pattern AB. The common terminating response of Patterns AA and BA was acquired faster than the discriminative terminating responses of Patterns BB and AB. The overall findings were consistent with the general notion of the importance of verbal mediating responses in concept formation and, more specifically, with predictions based on the four different patterns of relationships among initiating stimuli, mediating responses and stimuli, and terminating responses each in combination with more similar and less similar initiating stimuli.

REFERENCES

BAUM, N. H. A study in concept attainment and verbal learning. Unpublished doctoral dissertation, Yale University, 1951.

CAREY, J. E., & GOSS, A. E. The role of verbal labeling in the conceptual sorting behavior of children. *J. genet. Psychol.*, 1957, **90**, 69–74.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

FENN, J. D., & GOSS, A. E. The role of mediating verbal responses in the conceptual sorting behavior of normals and schizophrenics. *J. genet. Psychol.*, 1957, **90**, 59–67.

GOSS, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62**, 20–31.

GOSS, A. E., & MOYLAN, M. D. Conceptual block-sorting as a function of type and degree of mastery of discriminative verbal responses. *J. genet. Psychol.*, 1958, **93**, 191–198.

LACEY, H. M., & GOSS, A. E. Conceptual block sorting as a function of type of assignment of verbal labels and strength of labeling responses. *J. genet. Psychol.*, 1959, **94**, 221–232.

MANDLER, G. Associative frequency and associative prepotency as measures of response to nonsense syllables. *Amer. J. Psychol.*, 1955, **68**, 662–665.

(Received May 13, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 122–125

ROLE OF INSTRUCTIONS IN TWO-CHOICE VERBAL CONDITIONING WITH CONTINGENT PARTIAL REINFORCEMENT[1]

JOHN KOEHLER, JR.[2]

Indianapolis Center of Indiana University

In two-choice verbal conditioning with contingent partial reinforcement, *E* arranges for environmental event E_1 to follow *S*'s A_1 response with a fixed probability π_1 and, similarly, event E_2 to follow *S*'s A_2 response with a fixed probability π_2. In this conditioning procedure *S* is not given the outcome of the alternative, nonpredicted response on the nonreinforced trials. As a consequence, he may assume that either this response, if made, would have been rewarded, or the trial is a blank one[3] where no reward event follows either response. It should be obvious, then, that these hypothetical outcomes assumed by *S* can be a source of uncontrolled, and differential, reinforcement on the A_1 and A_2 response tendencies.

The present study attempted to partially control *S*'s choice of outcomes with the use of preconditioning instructions which varied in the information about outcomes on the nonreinforced trials. One group of *S*s (Group A) was told that all nonreinforced trials were to be treated as indicating the nonpredicted response, if made, would have been rewarded. Another group (Group B) was instructed that these trials were sometimes wrong predictions (i.e., Group A's instructions) and on other occasions, blank trials. A third group (Group C) was given no information about the nonreinforced trials.

An interesting aspect about these instructions groups is that the order among their mean terminal A_1 response rates can be predicted with statistical learning theory. According to the theory, a blank trial has no effect on either A_1 or A_2 response tendencies (Atkinson, 1956) so that if all nonreinforced trials were treated as blank trials, the expected asymptotic level of A_1 responding would approximate $\frac{\pi_1}{\pi_1 + \pi_2}$ (Neimark, 1956). On the other hand, if all nonreinforced trials were considered as wrong predictions, the expected asymptote would be $\frac{1 - \pi_2}{2 - \pi_1 - \pi_2}$ (Estes, 1954). It also can be easily shown with simple algebra that for $\pi_1 > \pi_2$, the first asymptotic expression is larger than the second one for all values of π_1, and π_2, between zero and one. Now, since Group B is instructed with both outcomes of nonreinforced trials, its mean response rate should approach some limit lying between these two theoretical asymptotes. The same would also apply to Group C inasmuch as the majority of its *S*s would be expected to treat some nonreinforced trials as blanks and others as wrong predictions even though not specifically instructed to do so. However for Group A, which is given only the wrong-prediction instructions, the terminal rate should approach the value of the second asymptotic expression and thus be lower than the other groups.

---

[1] This research was supported by a Faculty Research Grant from the Graduate School of Indiana University. The author is indebted to George Sumner, United States Army, for making available army trainees to serve as *S*s and for his work in a preliminary investigation of the problem treated here.

[2] Now at Chatham College, Pittsburgh, Pennsylvania.

[3] In the experimental literature, the blank trial is most frequently treated as an experimental operation where no response receives event E on a particular trial. In this study, however, it will refer only to *S*'s covert response of "blank trial" to a nonreinforced trial.

Previous attempts (Anderson & Grant, 1957; Brand, Sakoda, & Woods, 1957b) to influence response rates by varying instructions in the Humphreys-type verbal conditioning situation have met with failure, which may simply signify that the instructions employed in these studies were inadequate in producing the desired sets (Kendler, 1959, p. 69). The present study uses instructions that are somewhat different from those of the earlier studies and has the additional merit of being able to tie them to specific predictions from statistical learning theory.

An additional problem considered here was concerned with the value π_1 may take. Some investigators (Brand, Sakoda, & Woods, 1957a) have found evidence to question the assumed correspondence between instructions like those given to Group A and *S*'s tendency to treat the nonreinforced trials as instructed, particularly when the reinforcement levels are low. In the present study π_1 was factorized with the three sets of instructions while π_2 was held constant. In this case, however, statistical learning theory allows no specific predictions about the possible interaction between these variables.

METHOD

Subjects.—The *S*s were 120 United States Army trainees of college age at Ft. Benjamin Harrison. The *S*s reported in alphabetical order for the training session. Each then was randomly assigned to one of the experimental conditions but with the restriction that a replication of the design was completed before beginning with the next replication. Three *S*s were discarded but replaced with others, two because of *E* error and a third because he admittedly alternated his key presses.

Apparatus.—During training, *S* sat under a hood facing a stimulus panel: all *S*-side surfaces were painted flat black. At the base of the stimulus panel were mounted two telegraph keys and a pilot light above each key which served as the reinforcing event, E_1 or E_2. Also mounted on the panel, at *S*'s eye level, was a neon indicating lamp which signaled *S* to select and press one of the two keys.

The *E* sat on the opposite side of the panel from *S*, recording *S*'s key presses and programing E_1 and E_2 with mercury switches. The durations of the signal light, the reinforcing event, and the intertrial interval were automatically controlled with three Hunter Decade timers wired in series.

The experimental room was dimly illuminated and outside noises were minimized by scheduling training sessions only in evening hours.

Procedure.—After *S* was seated in front of the stimulus panel, the instructions were read to him and his questions answered by paraphrasing the appropriate parts of the instructions. Immediately following, *S* was run through a series of 200 uninterrupted conditioning trials. On each trial, the signal lamp was lighted for 1 sec. during which time *S* was to press one of the two keys; the appropriate E, or no E, followed in the second second; after the third second for the intertrial interval, the next trial began.

The instructions were the same for all three groups except for one section which was designed to influence *S*s' covert response to the nonreinforced trials. For Group A this section read as follows: "Sometimes the light above the key will come on. But on other occasions it will not, which means to you that you should have pressed the other key since its light was set to come on for that trial. You may not, however, press both keys. On every trial, you must always press one and only one key when the top light goes on." For Group B: "Sometimes . . . which means to you one of two things: Either, that you should have pressed the other key since its light was set to come on for that trial, or, that it didn't really matter which key you pressed because both lights were disconnected and it was impossible to light them by pressing the keys. You may not, however, guess that no light will appear." (The last sentence of the Group A instructions above.) For Group C: (The last sentence of the Group A instructions above.)

Design and randomized sequences of E.—Two levels of probability of reinforcement of A_1, $\pi_1 = .30$ or $.60$, were combined with the

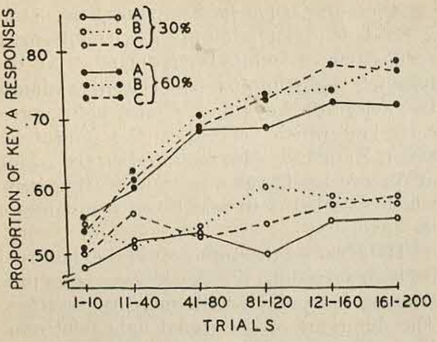


FIG. 1. The mean proportion of A_1 responses plotted as a function of trial blocks for six groups which received A, B, or C instructions and either 30% or 60% reinforcement on A_1.

three instructions groups to form a 3 × 2 design, having 20 *S*s in each cell.[4] For all six cells, π_2 was .10. A_1 was also counterbalanced on the right and left sides in each cell of the design.

The selection of key presses to be reinforced was done with 20 unrestricted randomized sequences, a different one for each *S* in a cell. The same 20 sequences were employed for the three 30% groups, and a similar procedure was followed for the 60% groups.

RESULTS AND DISCUSSION

The essential findings of the experiment are shown in Fig. 1 where the mean proportion of A_1 responses (Key A responses) are plotted in trial blocks. As predicted from statistical learning theory, the figure shows that the A groups have the lowest response rates over approximately the latter half of conditioning while the B and C groups are approaching similar rates near the close of the conditioning series.

The reliability of these differences was tested in a modified form of the analysis of variance. Using the A_1 response frequencies on the last 100 trials, the differences among the three instructions groups, combined on the π_1 variable, were tested against the within-groups variance estimate. Because the order among the terminal response rates of these groups was an a priori prediction, the following two orthogonal comparisons were tested rather than the overall comparison of group means: A vs. BC and B vs. C. The F ratio for the A vs. BC comparison was found to be 4.35 ($df = 1/114$, $.05 > P > .025$);[5] the F ratio for the B vs. C comparison was less than one.

A similar analysis also was made on the first 40 conditioning trials as a check for possible initial differences between the instructions groups: all the F ratios turned out to be less than one. However, these nonsignificant differences are not attributable necessarily to any larger variability in *S*s' response rates on these trials than on later ones. A comparison of the within-groups variance estimates of the first and last 40 training trials found them to differ but slightly (18.11 and 18.39, respectively), yet the reliability of differences among the three instructions groups on the last 40 trials was the same as reported above for the final 100 trials.

By contrast, the findings here stand as an exception to other studies (Anderson & Grant, 1957; Brand et al., 1957b) which failed to find any effect from varying instructions in the Humphreys-type verbal conditioning situation. As stated earlier, the instructions here differed from theirs, which may account for the difference in outcomes. Also, in view of our findings, the usual conclusion drawn that negative results with respect to instructions simply indicated the stability of behavior in this conditioning situation is now open to question.

---

[4] A seventh group of 20 *S*s received reinforcement probabilities, $\pi_1 = .60$ and $\pi_2 = .40$, and the Group C instructions. The purpose of this group was to find out if the procedure used here would result in matching between final response-levels and π values when $\pi_1 + \pi_2 = 1$ as previously reported (Neimark, 1956). A *t* test of the difference between the expected (.600) and observed (.553) proportions of A_1 responses for the last 40 trials, using the pooled within-groups error mean square as population variance estimator, gave a value of 1.93, which for 133 *df* lacks significance at the .05 level; the matching hypothesis, therefore, could not be rejected.

[5] In a recent article on multiple-comparison tests, Ryan (1959) has suggested (after Tukey) that where several comparisons between individual groups means are involved, the significance level be lowered to maintain a constant error rate. Because the a priori predictions dealt with in this study were based on a well-known theory, the .05 level was kept here.

An interaction between instructions and reward levels was expected on the basis that *S*s receiving the wrong-prediction instructions may disregard them when reward levels are low. Here, this means the 30%-A group should depart from its predicted asymptote more than should the matched 60% group. At first glance, Fig. 1 favors this possibility by the smaller difference in terminal response rates for the 30%-A vs. 30%-BC groups than for the similar 60% groups, particularly on the last trial block. On closer analysis, however, this interaction was found rather to result from the 30%-B and 30%-C groups having terminal rates relatively closer to 30%-A's predicted asymptote in comparison to the parallel relationship in 60% groups. Also to be noted is that failure of the A groups to depart from their predicted asymptotes in the expected direction was not consistent with the findings of others (Brand et al., 1957a). This discrepancy can be explained, in part at least, in terms of the several ways their instructions differed from Group A's instructions, even though they were logically equivalent to the latter. This would seem to be a reasonable assumption on the grounds that instructions in both studies were designed to influence how the *S* should consider the nonreinforced trial which, in turn, would affect response rates.

SUMMARY

The present study varied instructions which dealt with how *S*s should consider the nonreinforced trials in the two-choice contingent partial reinforcement situation: Group A was told to treat nonreinforced trials as making wrong predictions; Group B was told that these trials could be either wrong predictions or blank trials; Group C was told nothing about the nonreinforced trials. The order among the mean terminal response rates of the three instructions groups was predicted from statistical learning theory and confirmed: Group A had a lower mean rate during the last 100 of 200 trials than the other groups, and Groups B and C had somewhat similar mean rates. By contrast, differences among these groups on the first 40 conditioning trials were negligible.

Also investigated was a possible interaction effect between these instructions groups and reinforcement levels (30% and 60%) for the A_1 response. No interaction in the expected direction was obtained.

REFERENCES

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.

ATKINSON, R. C. An analysis of the effect of nonreinforced trials in terms of statistical learning theory. *J. exp. Psychol.*, 1956, **52**, 28–32.

BRAND, H., SAKODA, J. M., & WOODS, P. J. Contingent partial reinforcement and the anticipation of correct alternatives. *J. exp. Psychol.*, 1957, **53**, 417–424. (a)

BRAND, H., SAKODA, J. M., & WOODS, P. J. Effects of a random versus pattern instructional set in a contingent partial reinforcement situation. *Psychol. Rep.*, 1957, **3**, 473–479. (b)

ESTES, W. K. Individual behavior in uncertain situations. In C. L. Coombs, R. M. Thrall, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, 1954.

KENDLER, H. H. Learning. *Annu. Rev. Psychol.*, 1959, **10**, 43–88.

NEIMARK, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56**, 26–47.

(Received May 17, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 126–137

INFORMATION DEPRIVATION AS A MOTIVATIONAL VARIABLE[1]

AUSTIN JONES, H. JEAN WILKINSON, AND INA BRADEN

University of Pittsburgh

In the eleven years since psychologists were introduced to information theory by Miller and Frick (1949), several hundred articles in that area have appeared in psychological and related journals. In general, however, these studies have been tied closely (conceptually) to the electrical communication system model, and few attempts have been made to relate information theory to existing theoretical positions in psychology. Berlyne (1957b) has commented on the "unmistakable affinities" between information theory and S-R theory and suggested the value of attempting an integration of the two theories. In this context he proposed a "degree of conflict" measure (related to information measures of "uncertainty") and discussed its usefulness in assessing such variables as emotional disturbance, reaction time, drive, curiosity, stimulus complexity, and reward.

The research reported here is the first of a projected series intended to assess the relationship between information theory extensions in psychology and S-R reinforcement theory. This paper attempts to demonstrate that information deprivation functions as a motivational variable and that the information-deprived *S* will execute, with increasing frequency, instrumental responses which cause increasing amounts of information to be introduced into the environment.

Information is defined as the reduction of uncertainty in a receiver (Shannon, 1948). In order to deprive *S* of information, then, an experimental environment of maximal certainty of stimulation must be devised. In the limiting case, if for a given *S* there is perfect and continuing certainty regarding the aspects of the environment, then no uncertainty can be reduced. In the present study, maximal certainty of the environment, or information deprivation, was produced by isolating *S* in a lightproof chamber with audition greatly restricted. The environment is similar to that employed in various "sensory deprivation" studies (e.g., Bexton, Heron, & Scott, 1954; Heron, Doane, & Scott, 1956; Vernon & Hoffman, 1956), but serves here a very different theoretical purpose and is associated with a different dependent variable. In the present study the dependent variable was frequency of an instrumental response which presented patterned series of faint lights to *S*'s view.

The main hypotheses may be stated as follows: (*a*) Human *S*s in a condition of information deprivation will learn and maintain an instrumental response associated with the occurrence of light stimulation. (*b*) The rate of instrumental response will be an increasing monotonic function of hours of deprivation. (*c*) The frequency of instrumental response will be an increasing monotonic function of the amount of information contained in the patterning of the light stimuli.

---

[1] This study was supported in part by Grant M-2479 from the National Institute of Mental Health, United States Public Health Service. The junior authors were associated principally with the planning and execution of Exp. I. The assistance of Jo-Ann Gardner and John South in the execution of Exp. II is gratefully acknowledged. The authors are indebted to A. W. Bendig for statistical advice, to Charles Stelter for the design of programing apparatus, and to Harry Fowler for a critical reading of the manuscript.

The published research most clearly related to these hypotheses comes from the areas of perceptual "curiosity" and exploratory behavior. In a study of conflict and information theory variables related to human perceptual curiosity, Berlyne (1957a) utilized visual incentives differing in degree of relative uncertainty. The *S*s were allowed to respond as often as they wished to produce brief exposures of cards showing various patternings of simple geometric figures. Frequency of response was directly related to the asymmetry, irregularity, or complexity of figures, and inversely related to their redundancy. To the best of the writers' knowledge, this is the only study of an information dimension of visual incentives with human *S*s. Studies related to an information drive variable in human *S*s appear to be totally lacking.

EXPERIMENT I

Method

Subjects.—Eight undergraduate male volunteers from the University of Pittsburgh served in this study. They were paid $12.50 per day. Data relevant to a particular aspect of the problem are available for an additional 10 *S*s.

Experimental design.—The *S*s were maintained in a condition of information deprivation for a period of 12 hr. on each of 4 days. The four experimental sessions were separated by 2-day intervals. Four degrees of information in light series were employed; the relative information values were .00, .33, .67, and 1.00. On each day, *S*'s instrumental response was followed by one only of the four series. All *S*s received all four information series over the four days. The order in which *S*s received the information series over the four days was systematically varied in a latin square design.

Response functions plotted over the 12-hr. experimental session would be difficult to interpret if *S*'s fluctuations between sleep and wakefulness were not known. For this reason, the decision was made to try to keep all *S*s awake throughout by the administration of a 15-mgm. dexedrine spansule. This dosage was sufficiently low that no side effects of any sort were observed by *E*s, or reported by *S*s.

The information variable.—The *S*'s instrumental response consisted of a button press. Each response activated electronic programing apparatus which caused a series of light flashes to appear on a panel in the ceiling of the experimental room. There were two classes of light flashes, red and green. Each was produced by a very weak current to a small neon bulb behind a heavy glass lens. The two lenses were 4 in. apart. The flash was sufficiently faint as to illuminate no part of the room to the dark-adapted eye.

Each series, regardless of information value, always consisted of 24 flashes appearing singly at 1-sec. intervals. The apparatus was devised so that a light series in progress would not be affected by additional button presses. The various relative information values depended upon the proportion of the 24 flashes which were random (or "uncertain," from *S*'s standpoint). The relative information value of .00 was represented by zero randomness, i.e., 24 lights of a single, fixed color. The relative information value of 1.00 was represented by a random determination of color for each of the 24 flashes. The intervening values of .33 and .67 were associated, respectively, with one-third and with two-thirds of the 24 flashes determined randomly with regard to color. The latter two types of series are summarized as follows, where F refers to a flash of fixed color, R to a flash of randomly determined color:

I_{rel} = .33 FFRFFR etc.
I_{rel} = .67 FRRFRR etc.

The color of F, the fixed flash, was systematically varied over *S*s and over the four days and four information values. For a given *S* on a given day, however, the color of the fixed flash was always the same, regardless of the number of responses made.

Procedure.—Subjects were called in groups of three (the number of *S*s run concomitantly) to a brief meeting two or three days before the first experimental session. At that time they were instructed not to drink alcoholic beverages, and to go to bed by 11:00 P.M., on the evening preceding an experimental day. They were instructed not to eat or drink anything prior to coming to the experiment, as they would be given breakfast upon their arrival. Conditions of payment were also discussed. Total payment was to be made at the conclusion of Day 4; *S*s failing to appear for any of the four sessions would forfeit all pay.

Each experimental session began at 8:00 A.M. with *S*s to appear at 7:15. On Day 1, the three *S*s were given general instructions orally, in a group meeting. The experiment was introduced as a "study of boredom." The *S*s were informed that they were simply to lie quietly on a bed in a darkened room and that no activity of any kind would be permitted other than eating and using the toilet. An exception to this rule was that, whenever they wished, they could press a button at the side of the bed which would cause a series of light flashes to appear in the ceiling. Subjects were urged to avoid formulating rules governing their use of the button, and to respond or not simply as their moment-to-moment impulses directed. Complete instructions were given regarding such mechanical aspects of the experiment as the method of feeding, use of the toilet, etc.

The *S*s were then taken to separate, light-proof experimental rooms to await breakfast. Each room was bare of all furnishings except a bed, mounted to the floor on heavy felt pads. Subjects were instructed to practice going from the bed to the toilet with eyes closed, so that they could later do so in the dark without danger of falling.

After *S* had eaten breakfast, *E* entered to give final instructions. The *S* was shown the button on the end of a flexible cord at the side of the bed and informed exactly as to the nature of the light series that would occur following each press of the button. It was emphasized that the pattern of lights (whichever of the four information series was assigned to *S* on that day) would not change throughout the course of the day, and that *S* should feel free to use the button as often or as seldom as he wished. In explaining what kind of light series *S* would receive, *E* pointed out the serial location of the fixed and random flashes and made certain that *S* understood the concept of randomness and that a random flash was "rerandomized" in the series produced by further button presses. The *S*s were not informed, at the beginning of any of the four sessions, as to the nature of the light series to be encountered in subsequent sessions. At the conclusion of these instructions, *S* fitted beeswax and cotton plugs in his ears and put on acoustical "ear-muffs" (a product of Mine Safety Co. called "Noise-Foe"). The light was then removed from the room and *S* immediately made two "trial" responses, causing two series of 24 lights to appear. The beginning of the 12-hr. session was dated from that point; the two trial responses were not included in the data. A continuous recording of the button-pressing response was kept by Esterline-Angus equipment.

The *S*s were given no information as to the number or time of feedings they would receive. All *S*s received three meals, spaced 4 hr. apart, during each 12-hr. session. The time of the first meal, however, was varied systematically over the first 4 hr. of the session. The meals were put in trays which *E* pushed across the floor to the bed with a pole, while standing in the doorway of the experimental room. The doorway was "light-locked" to preserve the deprivation situation. The *S*s typically reported being unaware of the presence of *E* until the signal for food was given (two raps of the pole on the leg of the bed). At no other time was the experimental room entered.

At the conclusion of each 12-hr. session, *S*s individually answered a brief questionnaire concerning their reactions to the experiment and their physical comfort. This was principally for the guidance of *E*s and was not intended as an experimental variable.

After Day 1, *S*s went to their rooms immediately upon arrival, without repetition of the group-administered instructions of the first day. At the beginning of each session, however, the nature of the light series for that day was carefully explained individually to each *S*.

Results

Data relevant to the hypothesis that *S*s in a condition of information deprivation will learn and maintain an instrumental response associated with the occurrence of light stimulation, are available for the eight *S*s of the latin square design and for 10 additional pilot *S*s. Of these 18 *S*s, 14 made more than two responses in each experimental session. (Lower response frequencies were disregarded as possibly due to "chance" or non-intentional pressing of the button.)

For the eight *S*s of the latin square design, the mean numbers of responses for each of the four experimental days were, successively, 143, 42, 41, and 72. There was great inter-*S* variability; the *SD*s were 111, 43, 36, and 100. The 10 additional pilot *S*s did not each complete all four days,

TABLE 1

Analysis of Variance of Proportions of Responses: Exp. I

Source	*df*	*MS*	*F*
Days (D)	3	.035	3.50*
Information values (I)	3	.007	.70
Pooled error	18	.010	
residual	6	.017	2.83
D × *S*s within orders	12	.006	
Hours (3 hr. periods) (H)	3	.027	13.50***
linear	1	.040	20.00***
quadratic	1	.041	41.00***
cubic	1	.001	—
Pooled error	21	.002	
linear	7	.002	
quadratic	7	.001	
cubic	7	.004	
D × H	9	.006	3.00*
I × H	9	.003	1.50
Pooled error	54	.003	
residual	18	.005	2.50*
D × H × *S*s within orders	36	.002	

* $P < .05$.
*** $P < .001$.

and for that reason their data were not pooled with the others in the means and *SD*s just cited; their response frequencies were generally consistent with those values, however.

Table 1 summarizes the repeated measurements analysis of variance of trends (Grant, 1956) of the data for the eight *S*s of the latin square design. Data are the proportions of total responses tabulated over 3-hr. periods of each of the four experimental days; thus, for each *S* there are 12 values summing to one. Response was found to have both a significant linear and quadratic relationship to hours of deprivation (linear $F = 20.00$, $df = 1/7$, $P < .01$; quadratic $F = 41.00$, $df = 1/7$, $P < .01$). As may be seen in Fig. 1, the overall linear trend is an increasing one over time; the quadratic trend refers to the falling off of response in the final 3-hr. period. The distribution of responses over the four experimental days is shown in Fig. 2. Response is a significant decreasing function of days ($F = 3.56$, $df = 3/18$, $P < .05$). The Hours × Days interaction is also significant ($F = 2.81$, $df = 9/36$, $P < .05$), reflecting the fact that the response functions increase less rapidly over time upon successive experimental days. (Figure 2 shows an alternative plot of the data in which Subject AD was excluded. During Day 4 AD reported seeing a spark or speck of light inside his button. Curiosity over this caused him to press the button several hundred times in rapid succession, thus substantially raising the group mean for that day. The *E*s afterward could determine no malfunction of the button and were unable to verify *S*'s observation. Such an extended series of responses is highly atypical, and the alternative plot in Fig. 2 is offered to suggest that the "true" group function probably approaches asymptote on Day 4. Data for AD were included in the analysis of variance.)

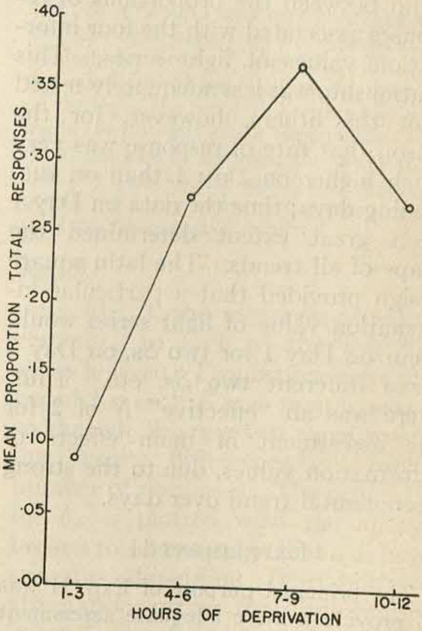


Fig. 1. Mean proportions of total responses as a function of hours of deprivation: Exp. I ($N = 32$; 8 *S*s pooled over 4 days).

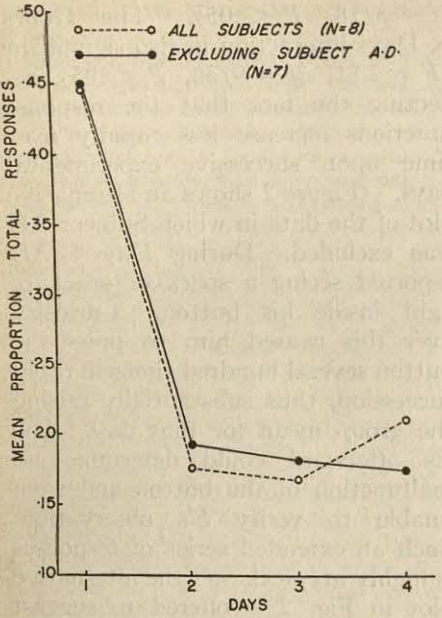


FIG. 2. Mean proportions of total responses as a function of successive days of experimentation: Exp. I.

No significant differences were found between the proportions of responses associated with the four information values of light series. This relationship was less adequately tested than the others, however, for the reason that rate of response was very much higher on Day 1 than on succeeding days; thus the data on Day 1 to a great extent determined the shape of all trends. The latin square design provided that a particular information value of light series would occur on Day 1 for two *S*s, on Day 2 for a different two *S*s, etc. Thus, there was an "effective" *N* of 2 for the assessment of main effects of information values, due to the strong decremental trend over days.

EXPERIMENT II

The principal purpose of Exp. II was to provide a more adequate assessment of the hypothesis that frequency of instrumental response will be an increasing monotonic function of the amount of information contained in the patterning of the light stimuli.

A second purpose was related to the interpretation of the finding in Exp. I that the rate of the instrumental response was linearly related to hours of deprivation. This suggests that information deprivation may function as do some physiological drives, in becoming more intense—and thus leading to increased goal activity—over increasing periods of deprivation. In the Hullian view, this interpretation is clouded by the formulation that performance is a function of drive and habit strength: $_sE_R = f(H \times D)$. In Exp I, all *S*s were free to respond and to receive the information incentive from the beginning of each session; thus, the increasing response function over time could be mainly or entirely a result of increased habit strength associated with increasing numbers of reinforced trials—without any contribution of an information drive variable. In Exp. II an attempt was made to examine the drive variable directly through the use of groups of *S*s differing in their degree of deprivation before being allowed access to the incentive.

A third purpose was to introduce a control series of light stimuli. Should it be found that *S*s respond more frequently to produce light series of higher degrees of information, it might be argued that the determining characteristic of the incentive is its variability rather than its uncertainty—or randomness—per se. For this reason, Exp. II included a fifth light series, which had zero relative information value, because the series was completely "fixed" or predictable, but which had maximum stimulus variability in the sense of maximum alternation of stimulus categories. This control series consisted of a fixed, single alternation of red and green flashes. If variability, or stimulus change, is a crucial aspect of the incentive, then *S*s should respond just as frequently to this control series, designated 0_A (zero information, alternation), as to the information series 1, which represents maximum information and also a high degree (random alternation) of stimulus change. If, however, it is information or uncertainty per se which is the crucial aspect of the incentive, then response to the 0_A series should not differ significantly from response to the 0 series, both being less frequent than response to series of higher information values.

Method

Subjects.—The *S*s were 26 undergraduate male students from the University of Pittsburgh. Each *S* was paid $12.00.

Experimental design.—Each *S* was maintained in a condition of information deprivation for one 10-hr. period. Subjects were divided into two groups of 13 each. The first group was not allowed to use the button or to receive the light series until a signal had been given 1 hr. after the beginning of the session. The other group was not allowed to use the button until a signal was given after 5 hr. had elapsed. In all other respects the groups were treated identically. The 1-hr. signal group is called Group I, and the 5-hr. signal group is called Group V. Not all *S*s successfully completed the experimental session; the causes of attrition will be noted under the appropriate sections below.

Each *S* was provided with a small control panel with one button and a dial which could be put in one of five positions. The *S* was permitted a "cafeteria" selection from among the the four light series of Exp. I, with relative information values of 0, .33, .67, and 1, plus the control series, 0_A. The order with which the various light series were assigned to dial positions was systematically varied. As in Exp. I, the color of the redundant or "fixed" flashes was constant within *S*s but systematically varied among *S*s.

Procedure.—Except as specifically indicated to the contrary, the procedure was in all ways identical to that of Exp. I. The *S*s arrived at 8:20 A.M. and received breakfast and instructions prior to entering the deprivation situation at 9:00 A.M. They were given group instructions similar to those for Exp. I, but with the following additional paragraph:

> It has been found in previous research that such boredom may be alleviated somewhat by the presentation of series of brief light flashes in the room. The purpose of this experiment is to find out what kinds of series work best—that is, what kinds of patterns of lights are most helpful in relieving boredom. As you lie on the bed in the darkened room, there is one thing you will be allowed to do. Whenever you wish, you may push a button at the side of the bed which will result in a series of brief light flashes in the ceiling. There will be a dial which may be set in any one of five positions, which will determine the kind of series—or pattern—which will occur. Exactly what the different series will look like you will see later. The main thing to remember is that you may push the button as often as you wish, for whichever pattern you wish.

At the conclusion of the instructions, *S*s were taken to their individual experimental rooms for breakfast and for final instructions. As in Exp. I, each *S* was given detailed instructions in the use of the apparatus and complete information as to the light series associated with each dial position. They understood that the light series associated with particular dial positions would not change in any way throughout the session. They were told not to use the apparatus until a signal was given, and that the time of the signal would be determined randomly. Following the signal, the apparatus would be continuously operative until the end of the session. Each *S* was instructed to check his understanding of the apparatus by trying each of the five light series twice in succession, following which the 10-hr. deprivation period was begun.

During the 10-hr. period, *S*s were fed in the manner described above for Exp. I. After 1 hr., *S*s of Group I were each given the signal that the light apparatus was available, and similarly for Group V after 5 hr. The signal was administered by *E* giving six short raps on the leg of the bed with a pole while standing in the doorway of the experimental room. As in the case of feedings, the doorway was carefully light-locked and *S*s afterward reported not being aware of the presence of *E* until the raps actually occurred.

Results

In Fig. 3, the mean proportion of responses to each of the five light series is given. Proportions were computed for each *S* and then averaged, so that all *S*s are given equal weighting despite differences in absolute number of responses. The proportion for 0_A is plotted with the abscissa broken to acknowledge that a different stimulus dimension (variability or alternation, rather than information) is involved. Table 2 summarizes the analysis of variance of trends for these data (Group I and V combined)

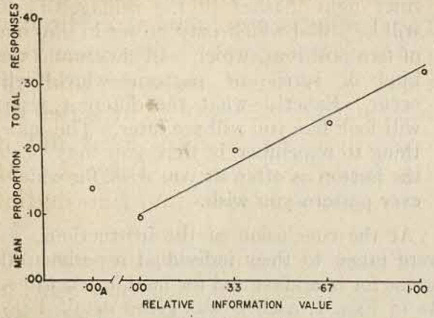


FIG. 3. Mean proportions of responses associated with the five light series: Exp. II.

excluding the values for 0_A, however, for the reason just noted, that a different stimulus dimension is involved. Proportion of response is shown to be a highly significant linear function of the relative information value of the light series ($F = 23.33$, $df = 1/25$, $P < .001$). No significant differences were found between proportions of response to 0 and 0_A or between 0_A and .33 ($P > .05$ in each case).

In Fig. 4, the mean proportions of responses are plotted individually for Groups I and V as a function of hours of deprivation. Separate analyses of variance of trends were performed, over the 9 hr. of responding for Group I and over the 5 hr. of responding for Group V. Interaction terms were computed according to the method suggested by Myers (1959). The analysis for Group I is summarized in Table 3. The data are the proportions of responses made to each of the four information values, 0, .33, .67, and 1, in each of the nine 1-hr. periods. Thus, for each *S* the data consist of 36 proportions summing to one. The analysis is based upon 11 rather than 13 *S*s; the data for two *S*s are incomplete owing to apparatus malfunction.

TABLE 2

ANALYSIS OF VARIANCE OF PROPORTIONS OF RESPONSES ACCORDING TO THE INFORMATION VALUES OF LIGHT SERIES: EXP. II

Source	*df*	*MS*	*F*
Information values (I)	3	.24	
linear	1	.70	23.33***
quadratic	1	.00	
cubic	1	.01	
Between *S*s	25	.04	
I × *S*s	75	.03	
linear	25	.03	
quadratic	25	.04	
cubic	25	.03	

*** $P < .001$.

Response shows a significant downward linear trend over hours ($F = 5.53$, $df = 1/10$, $P < .05$). There was a significant linear relationship between response and the information values of light series ($F = 23.04$, $df = 1/10$, $P < .001$), which is consistent with the analysis summarized in Table 2, based upon pooled data of Groups I and V. All interactions of Hours and Information Values were nonsignificant, except for the cubic Information ×quartic Hours component ($F = 5.02$, $df = 1/10$, $P < .05$). Due to the relatively large number of simultaneous interaction *F*s, and to the absence of any clear rationale for such an interaction, this finding is interpreted as due to chance fluctuations in the data.

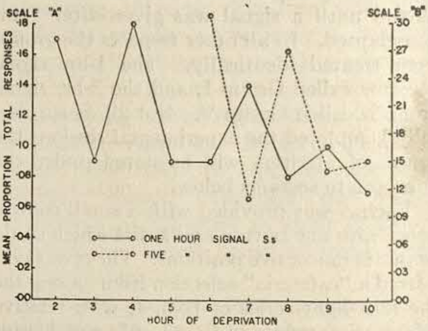


FIG. 4. Mean proportions of total responses as a function of hours of deprivation: Exp. II. (For ease of comparison, Group I is plotted on Scale A, Group V on Scale B.)

Table 4 summarizes a similar analysis of variance for the data of Group V. The analysis is based upon the

TABLE 3

Analysis of Variance of Proportions of Responses for Group I: Exp. II

Source	*df*	*MS*	*F*
*S*s	10	.0000	
Hours (H)	8	.0040	1.29
linear	1	.0083	5.53*
quadratic	1	.0001	—
cubic	1	.0007	—
quartic	1	.0000	—
quintic	1	.0000	—
sextic	1	.0034	—
septic	1	.0191	3.35
octic	1	.0000	
H × *S*s	80	.0031	
linear	10	.0015	
quadratic	10	.0047	
cubic	10	.0017	
quartic	10	.0014	
quintic	10	.0016	
sextic	10	.0042	
septic	10	.0057	
octic	10	.0042	
Information values (I)	3	.0192	4.36*
linear	1	.0553	23.04***
quadratic	1	.0005	—
cubic	1	.0018	—
I × *S*s	30	.0044	
linear	10	.0024	
quadratic	10	.0076	
cubic	10	.0033	
I × H[a]	24	.0016	
I × H × *S*s	240	.0028	

[a] All components of interaction also nonsignificant, with single exception noted in text.

* $P < .05$.

*** $P < .001$.

eight *S*s for whom no apparatus malfunction occurred and who did not violate the requirements of the experiment by "jumping the gun" and responding before receiving the signal. (That the latter was possible at all was due to an unforeseen deficiency in the apparatus which was subsequently remedied.) As in the analysis for Group I, the data are the proportions of responses associated with each information value in each of the 5 hr. of responding. The trend analysis over hours shows both the quadratic and quartic trends to be significant (quadratic $F = 50.29$, $P < .01$; quartic $F = 6.95$, $P < .05$; both with 1/7 *df*). The quadratic trend indicates a negatively decelerated rate of response over time. The quartic trend reflects a somewhat cyclic fluctuation of response. The linear relationship between response and the information values of light series did not attain significance, as it did in the pooled analysis of Table 2. All interactions of Hours and Information Values were nonsignificant.

Figure 5 compares the data of Groups I and V in their first 5 hr. of access to the incentive. Data are the mean absolute numbers of responses during each hour. In order to provide equal *N*s for the analysis of variance of trends (summarized in Table 5), the first eight *S*s run in Group I were compared to the eight *S*s of Group V. The finding of principal interest regarding the interpretation of information deprivation as drive is the significant quadratic component of the Hours × Groups interaction ($F=7.52$, $df=1/14$, $P<.02$). This interaction refers to the fact

TABLE 4

Analysis of Variance of Proportions of Responses for Group V: Exp. II

Source	*df*	*MS*	*F*
*S*s	7	.0000	
Hours (H)	4	.0171	6.11**
linear	1	.0128	2.13
quadratic	1	.0352	50.29***
cubic	1	.0071	2.96
quartic	1	.0132	6.95*
H × *S*s	28	.0028	
linear	7	.0060	
quadratic	7	.0007	
cubic	7	.0024	
quartic	7	.0019	
Information values (I)	3	.0163	1.79
linear	1	.0396	2.61
quadratic	1	.0039	.39
cubic	1	.0056	2.80
I × *S*s	21	.0091	
linear	7	.0152	
quadratic	7	.0100	
cubic	7	.0020	
I × H[a]	12	.0021	1.24
I × H × *S*s	84	.0017	

[a] All components of interaction also nonsignificant.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

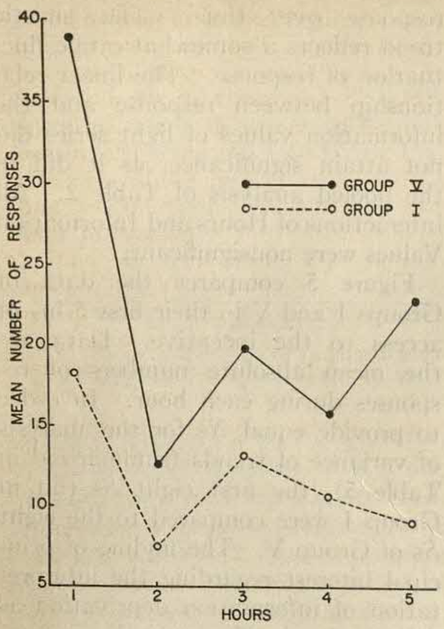


FIG. 5. Mean absolute numbers of responses during the first 5 hr. of access to the incentive: Exp. II.

that, in their first 5 hr. of access to the incentive, *S*s of Group V began by responding at a much higher rate than *S*s of Group I, but in successive hours responded at a lesser rate, approximating that of Group I, finally increasing their rate differential again in the fifth hour. The significant linear and quadratic trends over hours duplicate the finding of the earlier analyses (Tables 3 and 4). Significant individual differences were demonstrated.

TABLE 5

ANALYSIS OF VARIANCE OF ABSOLUTE NUMBER OF RESPONSES FOR GROUPS I AND V OVER FIRST 5 HR. OF RESPONDING: EXP. II

Source	*df*	*MS*	*F*
Hours (H)	4	791	7.26***
linear	1	792	8.16*
quadratic	1	1216	15.79**
residual	2	578	
Groups (G)	1	2112	2.47
H × G	4	183	1.69
linear	1	76	—
quadratic	1	579	7.52*
residual	2	42	
*S*s	14	854	7.83***
*S*s × H	56	109	
linear	14	97	
quadratic	14	77	
residual	14	132	

* $P < .05$.
** $P < .01$.
*** $P < .001$.

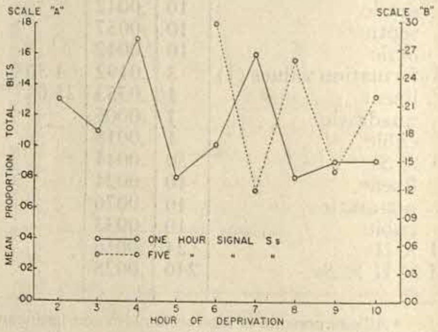


FIG. 6. Mean proportions of total bits received by *S*s as a function of hours of deprivation: Exp. II. (For ease of comparison, Group I is plotted on Scale A, Group V on Scale B.)

The analyses of variance reported thus far, except for the last, have been based upon proportions of total responses as the dependent variable. It was felt that this was the most meaningful dependent variable, though the choice was to some extent an arbitrary one. The possibility remained that some other dependent variable might have yielded additional results of importance. As a check on this point, the data were retabulated in terms of the proportions of total bits of information received by *S* which were associated with the various information values and hours of deprivation. These data are presented graphically in Fig. 6. A comparison of Fig. 4 and 6 shows a clear congruence between the plots of the two different dependent variables. The analyses of variance of trends for Groups I and V were repeated employing the variable of proportion of bits, with the same results as in the previous analyses (Tables 3 and 4).

DISCUSSION

Information Deprivation as Drive

The combined results of Exp. I and II support the view that information deprivation functions as a drive variable in the same sense as the well-studied homeostatic drives of hunger, pain, and thirst.

An adequate test of information deprivation as a motivational variable rests, first of all, upon *E*'s success in eliminating all "demand" characteristics of the experimental situation for each *S*, a problem recently discussed in another context by Orne (1959). That is, *S* must feel entirely free of implicit social pressure to perform the response expected by *E*. In this experiment *E*s took great care to impress upon each *S* that to perform in accordance with any conception he might have of *E*'s views was a violation of the requirements of the experiment, that *E* wished only that *S* would do as he himself felt. Postexperimental questionnaire and interview data indicate that this effort was successful for most *S*s. For three of the four *S*s (out of 18) in Exp. I who made less than the criterion of two responses, it was subsequently discovered that they had persisted in viewing the experiment as a test of their personal strength, in which making the instrumental response was an admission of weakness. (The second experiment was more successful in eliminating this problem.) Fourteen of the 18 *S*s of Exp. I did perform in excess of the criterion number of responses; thus for the majority of *S*s, information deprivation clearly functioned as a motivational variable as determined by their learning and maintaining an instrumental response serving to introduce the relevant incentive.

In designing Exp. I, the attempt was made to select an incentive of very low magnitude so as to provide for an increasing response function over a relatively long period of deprivation. Too great an incentive would lead to early satiation, with a decrement of response and a less clear opportunity to observe the energizing effects of successive hours of deprivation. That the incentive, in fact, was of low magnitude is attested by Fig. 1; response rose steadily over the first 9 hr., falling off somewhat in the last 3 hr. of the session. The reason for that decline in response is not clear. There is some evidence that *S*s had minimal cues as to the passage of time and that they anticipated the end of the session. It is possible that for some *S*s the dexedrine spansule was no longer effective, so that they slept in the final hours, thus depressing the group response function. If the decrement is not an artifact, it might be inferred that the satiation part of a drive-satiation cycle had begun.

That information deprivation is a motivational variable seemed reasonably clear from Exp. I. That it functioned as *drive* in the Hullian formulation was, however, not clear owing to the fact that habit strength and drive were confounded as determinants of increased instrumental response over time. In Exp. II, *S*s of Group I were deprived of access to the incentive for 1 hr. and *S*s of Group V were similarly deprived for 5 hr. According to the information drive hypothesis, *S*s of Group V would be expected, in their first hour of access to the incentive, to make the instrumental response much more frequently than would the *S*s of Group I in a comparable period. Such a difference would be due to the differential effect of drive level on the performance of the two groups, since initial habit strength is presumably the same in each group. With successive hours of continuous access to the incentive, however, the effects of the difference in drive (hours of deprivation) should diminish, so that the high rate of response in Group V should decrease and more nearly approximate the response rate of Group I. Such a relationship did, in fact, occur. In general the data support the view that the increasing response rate over deprivation time observed in Exp. I (Fig. 1) is dependent at least in part upon increases in drive rather than increases in habit strength alone.

A finding that is somewhat troublesome to interpret is that of decreasing response over the 4 days of Exp. I (Fig. 2). This apparent "accommodation" effect may be the result of several factors which may only be speculated about at this stage. One of the more plausible explanations has to do with the presence of irrelevant drive, anxiety, which differentially increases performance on Day 1. The *S*s of Exp. I, committing themselves to a 4-day series, were understandably somewhat apprehensive on Day 1. After Day 1, when they learned that *E*s did not in fact stress or "trick" them in any way, it seems likely that their anxiety was sharply reduced, with the result that performance dropped to a level associated principally with the relevant drive—information deprivation.

Information as Incentive

The results of Exp. II clearly verify the hypothesis that the frequency of instrumental response is an increasing monotonic function of the amount of information contained in the light series. Also, the relationship is linear.

It is believed that the incentive was, specifically, a stimulus serving to reduce uncertainty in *S*, i.e., a stimulus containing information—or some degree of unpredictability. It was shown that when *S*s respond maximally to information series 1, they are not responding simply or primarily to the alternation of stimulus categories, for they respond at a relatively low rate when the series contains maximal alternation but is completely redundant. This is not to say that differences in degree of alternation or stimulus change may not have differential incentive value when relative information is held constant. From the present study, however, we infer at least that information is the more potent dimension of incentive.

For both Exp. I and II, the procedure was devised so as to reduce as nearly to zero as possible any "problem-solving" motivation underlying the response of button pressing. Had *S*s been given no knowledge of the nature of the light series, it seems likely that considerable button pressing might have occurred associated with their attempts to discover the patterning of the light flashes. Instead, *S*s of these experiments were given complete knowledge of the statistical characteristics of the light series, plus orientation trials in which *S* was able concretely to verify what *E* had told him. Thus, once the experimental session was begun, *S*s understood that their responses could not result in additional understanding of the patterning of the light series. With each button press, their uncertainty would be reduced only by the presentation of flashes whose color had been established randomly.

Information Concepts Applied to Exploratory and Curiosity Behavior

The results of this study suggest that information theory concepts may be helpful in achieving some unification of the various related, but somewhat unintegrated views of exploratory, manipulatory, and curiosity behavior. It seems possible that all such behaviors may be referred to information motivation or drive, i.e., motivation for the reduction of uncertainty. Whether the organism is motivated by manipulatory incentives (Harlow & McClearn, 1954), visual exploration (Butler, 1953), maze exploration (Montgomery, 1954; Montgomery & Segall, 1955), the investigation of novel stimulus objects (Berlyne, 1955), or the investigation of portions of the environment in which cue alternation has occurred (Kivy, Earl, & Walker, 1956; Williams & Kuchta, 1957), the reduction of uncertainty may be regarded as the crucial aspect of the behavior. A virtue of the information formulation is that it is completely general, and may be applied to any performance, to any stimulus situation.

The utility of the informational approach seems particularly clear in the quantification of the incentive, a problem apparently not dealt with previously except for a somewhat tangential treatment in Berlyne's (1957a) study of perceptual curiosity. How many units of incentive are associated with a mechanical puzzle? How many with a checkerboard maze or with a certain period of visual exploration? Formal information measures apply, in principle, to each of these situations. More importantly, information measures may be used in the construction of both unidimensional and multidimensional incentives which would permit more exact tests of many of the ideas underlying what is usually referred to as exploratory or curiosity behavior. Finally, a semantic contribution may be made by the information quantification of the incentive in that it refers to the probabilities of events as viewed by *S*, and may be clearly separated from the sheer physical fluctuation of stimulus categories. Heretofore both aspects of incentive have tended to be lumped together confusingly under such terms as novelty or stimulus variability.

SUMMARY

This study examined the hypotheses that information deprivation functions as a motivational variable and that information-deprived *S*s will execute, with increasing frequency, instrumental responses which cause increasing amounts of information to be introduced into the environment.

College *S*s served in a condition of information deprivation produced by isolation in lightproof chambers for 8- or 12-hr. periods. In Exp. I, they were allowed to press, whenever they wished, a button which caused patterned series of lights to appear in the ceiling. The *S*s received a different degree of relative information in the light series on each of 4 experimental days. Response was found to be an increasing linear function of hours of deprivation, with a slight decrement in response in the last quarter of the 8-hr. session. Response did not show a significant relationship to the information values of light series.

In Exp. II, designed to provide a more sensitive test of that relationship and to further explicate the motivational variable, *S*s were run over a single 10-hr. period, with a "cafeteria" selection of the light series made possible by a dial which could be readily manipulated in the dark. One group of *S*s was allowed access to the dial after a delay of 1 hr., a second group after 5 hr. Response was shown to be a highly significant increasing monotonic function of the information values of light series. The 1-hr. and 5-hr. groups were shown to have different trends in response over hours, which was interpreted as support for a drive formulation of information deprivation.

REFERENCES

BERLYNE, D. E. The arousal and satiation of perceptual curiosity in the rat. *J. comp. physiol. Psychol.*, 1955, **48,** 238–246.

BERLYNE, D. E. Conflict and information theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.*, 1957, **53,** 399–404. (a)

BERLYNE, D. E. Uncertainty and conflict: A point of contact between information-theory and behavior-theory concepts. *Psychol. Rev.*, 1957, **64,** 329–339. (b)

BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8,** 70–76.

BUTLER, R. A. Discrimination learning by rhesus monkeys to visual exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46,** 95–98.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53,** 141–154.

HARLOW, H. F., & MCCLEARN, G. E. Object discrimination learned by monkeys on the basis of manipulation motives. *J. comp. physiol. Psychol.*, 1954, **47,** 73–76.

HERON, W., DOANE, B. K., & SCOTT, T. H. Visual disturbances after prolonged perceptual isolation. *Canad. J. Psychol.*, 1956, **10,** 13–18.

KIVY, P. N., EARL, R. W., & WALKER, E. L. Stimulus context and satiation. *J. comp. physiol. Psychol.*, 1956, **49,** 90–92.

MILLER, G. A., & FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56,** 311–324.

MONTGOMERY, K. C. The role of the exploratory drive in learning. *J. comp. physiol. Psychol.*, 1954, **47,** 60–64.

MONTGOMERY, K. C., & SEGALL, M. Discrimination learning based upon the exploratory drive. *J. comp. physiol. Psychol.*, 1955, **48,** 225–228.

MYERS, J. L. On the interaction of two scaled variables. *Psychol. Bull.*, 1959, **56,** 384–391.

ORNE, M. T. The nature of hypnosis: Artifact and essence. *J. abnorm. soc. Psychol.*, 1959, **58,** 277–299.

SHANNON, C. E. A mathematical theory of communication. *Bell Sys. tech. J.*, 1948, **27,** 379–423, 623–656.

VERNON, J., & HOFFMAN, J. Effect of sensory deprivation on learning rate in human beings. *Science*, 1956, **123,** 1074–1075.

WILLIAMS, C. D., & KUCHTA, J. C. Exploratory behavior in two mazes with dissimilar alternatives. *J. comp. physiol. Psychol.*, 1957, **50,** 509–513.

(Received May 19, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 138–144

RESPONSE PROBABILITY IN A TWO-CHOICE LEARNING SITUATION WITH VARYING PROBABILITY OF REINFORCEMENT[1]

ROBERT H. HICKSON[2]

Indiana University

An asymptotic matching theorem from statistical learning theory states that the mean asymptotic probability of a response is equal to the probability that the response is reinforced (Estes & Straughan, 1954). This prediction has been confirmed for several experimental situations where a fixed probability of reinforcement has been employed for each response alternative (e.g., Brunswik, 1939; Estes & Straughan, 1954; Grant, Hake, & Hornseth, 1951; Lauer & Estes, 1954). Estes (1957) has recently extended the asymptotic matching theorem to include cases of variable probability of reinforcement. The purpose of the present study is to test this theory for the special case in which probability of reinforcement (π) approaches an asymptote exponentially. Intuitively, it would seem reasonable to suspect that as π changes for two or more response alternatives the organism will, more or less gradually, adapt his response stratagem to the change in π schedule. Furthermore, although we might predict asymptotic response probabilities equal to π and $1 - \pi$ in a two-choice situation, we should perhaps expect a "lag" in response probabilities behind the changing π and $1 - \pi$ values. These two notions of gradual adaptation to the changing π value, and a "lag" of response probabilities behind π are exactly predicted in Estes' (1957) theory.

Ramond (1954), using a two-choice bar-pressing apparatus similar to the one used in the present experiment, found that for rats on 22-hr. food deprivation the choice curve differed in form from typical selective learning curves in that it reached a maximum and then slowly declined. With fixed π values of .67 for one bar and .33 for the other bar, *S*s developed a significant preference for the .67 bar, and in fact reached response probabilities of greater than .80 to the more frequently reinforced bar before beginning to decline. As elaborated by Ramond (1954), Spence's model (1956) for the forced-choice situation predicts that response probability to the more frequently reinforced bar would, at maximum, exceed the probability of reinforcement before declining to approach chance. On the other hand, statistical learning theory predicts that response probability to the more frequently reinforced bar would approach π exponentially. Since there is no empirical evidence to support either theoretical model in either a forced-choice or correction situation where varying π is employed with rats, both the forced-choice and correction procedures were used in the present experiment.

The specific purposes of this study were: (*a*) to test two predictions of Estes' statistical learning theory for the special case of varying probability of reinforcement (π) in which π approaches an asymptote exponentially;

---

[1] This study is based upon a portion of a dissertation submitted to the Graduate School of Indiana University in partial fulfillment of the requirements for the PhD degree. The writer would like to express his appreciation to W. K. Estes, who directed the research. The experiment was conducted at Defence Research Medical Laboratories, Toronto, Canada.

[2] Now at University of Arkansas, Fayetteville.

viz., the predictions concerning learning rate and terminal response probability; (*b*) to compare response probabilities obtained under two different procedural conditions; viz., a correction procedure and a forced-choice procedure; and, (*c*) to test the effect upon subsequent learning rate and terminal response probability of two methods of establishing initial response probability; viz., "pretrained" initial response probability vs. "assumed" initial response probability.

METHOD

Subjects.—The *S*s were 120 female albino rats of the Wistar strain. All *S*s were naive and between 90 and 110 days of age at the beginning of training.

Apparatus.—Two pieces of apparatus were used. The first consisted of a startbox connected to a 12-in. runway. A bar-press mechanism and food dispenser were at the end of the runway opposite the startbox. The purpose of the apparatus was to establish a stable bar-press latency for each *S*.

The main apparatus, a two-choice bar-pressing chamber connected to a startbox, has been described in detail elsewhere (Ramond, 1954). Briefly, it consisted of a triangular response chamber 18 in. across at the rear, 12 in. deep, and 10 in. high. A startbox led into the front apex of the response chamber. The startbox, which was similar to the one used in the runway apparatus, was 4 in. high, 3 in. wide, and 8 in. long. In the two-choice chamber either, or both, of the two bars could be presented through holes placed near the edge of the walls extending from the door from the startbox. Raising the metal door of the startbox started two Standard Electric timers. By pressing a bar $\frac{1}{16}$ in. with a pressure of approximately 2 gm. *S* stopped one of the timers; by pressing the other bar *S* stopped the other timer. Through a relay circuit the depression of a bar activated a solenoid which simultaneously retracted the bar and, through a $\frac{1}{4}$-in. hole directly below the bar, presented a 45-mg. pellet of food in a Plexiglas dispenser. A shielded 7-w. bulb was located above the bar of the training runway and above each bar of the two-choice apparatus. The total illumination in the two-choice apparatus was such that each bar and food dispenser could not be seen by *E* (nor, presumably, by *S*) unless the light above was on.

Feeding schedule.—All *S*s were placed on a 22-hr. feeding schedule for seven days prior to preliminary training in the runway. Each day each *S* was given 8 gm. Purina chow mixed with an equal quantity of water and a small amount of cod liver oil. The *S*s were fed in individual cages for 30 min. during the same hour each day.

Runway training.—Subsequent to the 7 days of 22-hr. feeding schedule training, the 120 *S*s were given 16 days of bar-press training in the single bar runway. Each *S* was given three trials per day. The interval between trials during runway and two-choice training was 15 sec.

Pretraining in the two-choice apparatus.—For each *S* the dispensation of a food pellet from one of the dispensers in the two-choice apparatus was designated E_1, and from the other dispenser as E_2. To counterbalance for position tendencies 10 *S*s in each group were assigned E_1 to the right dispenser; the other 10 *S*s were assigned E_1 to the left dispenser. The probabilities of occurrence of the two reinforcing events, E_1 and E_2, were π and $1 - \pi$, respectively. The behaviors available to *S* were categorized into two classes, A_1 and A_2, the former being responses resulting in depression of the bar over the E_1 dispenser, and the latter responses resulting in depression of the bar over the E_2 dispenser.

Following the 16 days of runway training *S*s were randomly assigned to six groups of 20 *S*s each. During pretraining Group C1 (correction group) and Group F1 (forced-choice group) each received 100% of trials reinforced to E_2, and no reinforcement to E_1. Group C.5 (correction group) and F.5 (forced-choice group) each received 50% of trials reinforced to E_2, and 50% to E_1. Groups CN.5 and FN.5 received no pretraining in the two-choice apparatus, but were assumed to have equal response probabilities to each of the two bars.

The procedure employed with all correction groups for the daily training trials during both two-choice procedures was the following: *S* was placed in the startbox; 5 sec. after the startbox lid was closed the door leading into the two-choice chamber was opened; both lights were on and both bars were present. If *S* made an A_1 response on an E_1 trial *S* was reinforced for that response and approximately 3 sec. later was removed from the chamber. If *S* made an A_2 response on an E_1 trial the bar was withdrawn and *S* was permitted to "correct" his response; that is, *S* was permitted to make an A_1 response, which was reinforced. The same procedures were followed if S made an A_2 response on an E_2 trial (i.e., reinforced); or, an A_1 response on an E_2 trial (i.e., permitted to correct). Both lights were on for the whole time S was in the chamber.

The procedure employed with all forced-choice groups for the daily training trials during both two-choice procedures was the following: S was placed in the startbox; 5 sec. after the startbox lid was closed the door leading into the two-choice chamber was opened. On E_1 trials the corresponding light was on and the bar was present, whereas the light corresponding to E_2 was off and the A_2 bar was withdrawn. Similarly, on E_2 trials the corresponding light was on, and the A_2 bar was present whereas the light corresponding to E_1 was off and the A_1 bar was withdrawn from the chamber. The response choice of S, therefore, was forced according to the experimental selection of E_1 and E_2 trials. On all daily training trials for all Ss in the forced-choice groups only one light was on and only the bar corresponding to that light was present.

For each S response probabilities for each bar-press response were evaluated by preceding the three training trials of a given day with a test trial. On these single test trials, for all groups which were given pretraining, both lights were on and both bars were present. On all test trials each S was reinforced for either response, A_1 or A_2, and 3 sec. later was removed from the apparatus. Test trials were given on alternate days only. During pretraining test trials were given on Days 2, 4, 6, 8, and 10. The test trial was followed immediately by the three daily training trials.

Varying π procedure.—After establishing stable A_1 and A_2 response probabilities for the four groups which received pretraining in the two-choice apparatus, all six groups were given 30 days of training in the two-choice apparatus under conditions of varying π. For all groups π was varied according to the function $\pi = a + cb^n$, where $a = .67$, $c = -.20$, and $b = .85$. The reinforced side was established randomly for each trial according to the π value appropriate for that trial. No attempt was made to smooth the resulting π function for each S or each group. The Ss in Groups C1, C.5, and CN.5 (correction groups), and in Groups F1, F.5, and FN.5 (forced-choice groups) were each given three daily trials according to the methods previously prescribed for correction and forced-choice procedures. The method employed for evaluating response probabilities for each bar-press response during the varying π procedure was identical with the method employed during pretraining: beginning with Day 2, and on alternate successive days, a test trial preceded the three daily trials.

RESULTS

Runway latencies and θ estimates.—An analysis of variance showed no significant difference among groups in the asymptotic response speeds attained during runway training ($F = .51$, $df = 5/114$). The analysis was performed on the last six trials of runway training. It was assumed that any variation attributable to runway training which transferred to the two-choice situation was equally represented in all six experimental groups. The analysis showed a significant difference among trials ($F = 8.61$, $df = 5/570$) and it was concluded that the groups of Ss had not reached stable asymptotic response speeds.

A parameter θ was evaluated for each group by substituting the runway latency data into an equation described by Estes (1959, p. 394), and solving for θ. The equation relates mean latency to h (the minimum time required to execute a response; which in this experiment was the mean for each group of the shortest latency during runway training for each S for each block of three daily trials), and n, the number of reinforced trials. A θ value was calculated for each block of three daily trials and the mean θ value computed for the 16 blocks of trials for each group of Ss. The θ value obtained for each group was: C1 = .093; F1 = .079; C.5 = .067; F.5 = .101; CN.5 = .113; FN.5 = .064.

Two-choice pretraining.—The results of pretraining are shown in Fig. 1, which shows the percentage choice of A_1 responses on test trials. Five test trials were given during pretraining before it was decided that Groups C1 and F1 had reached the re-

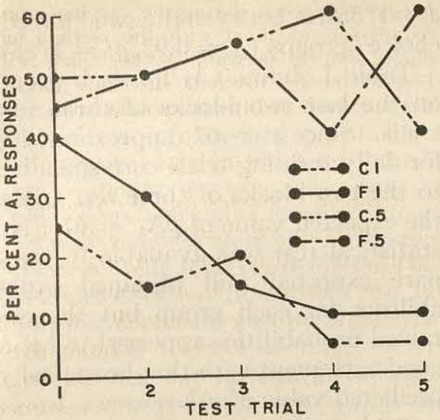


FIG. 1. Percentage choice of A_1 responses on test trials during pretraining.

quired response probability level of approximately zero for A_1 responses.

Variable π procedure.—The two-choice behavior for the six groups is shown in Fig. 2, which shows percentage choice of A_1 responses on test trials for the three correction groups and for the three forced-choice groups. It is apparent that *S*s in Group C1 gradually shifted from a small percentage of A_1 responses to a terminal level between 60 and 70%. Similarly, *S*s in Groups C.5 and CN.5 gradually shift percentage A_1 responses to the asymptotic probability of reinforcement. However, it is obvious that all forced-choice groups reach response probabilities far in excess of the maximum π value before eventually declining to approximate the appropriate π value.

Theoretical predictions.—An equation has been provided (Estes, 1957, Eq. 2C) which relates response probability to the number of trials, the slope parameter θ, and the constants employed in the variable π function. The adequacy of this theoretical function in predicting the performance of the correction and forced-choice groups on test trials is shown in Fig. 2. The three smooth functions shown in Fig. 2 describe the theoretically predicted percentage of A_1 responses during the varying π procedure for the three correction groups. The top function (the first to reach asymptote) describes the predicted learning rate and asymptotic response probability for Group CN.5. The middle function and lower function describe the predicted learning rate and asymptotic response probability for Groups C.5 and C1, respectively. It is clear from Fig. 2 that the predicted learning rates and terminal response probabilities are in close agreement with empirically obtained results for the three correction groups. Since the forced-choice groups shown in the lower part of Fig. 2 all reached response probabilities substantially above the maximum level (.67) predicted by the theory no detailed

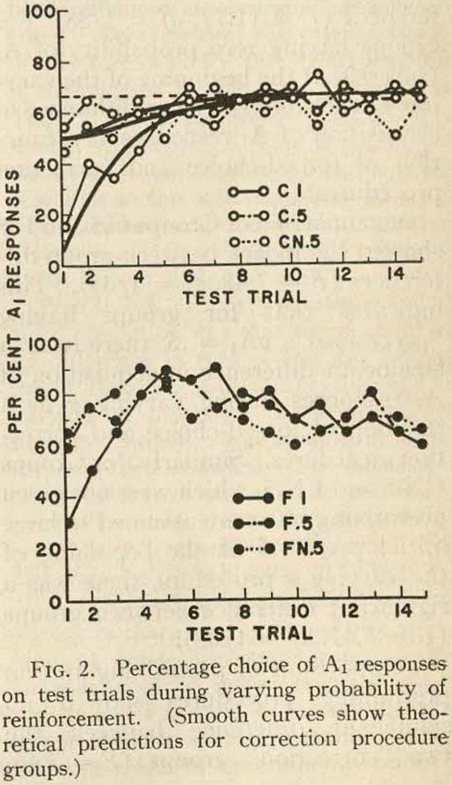


FIG. 2. Percentage choice of A_1 responses on test trials during varying probability of reinforcement. (Smooth curves show theoretical predictions for correction procedure groups.)

TABLE 1

MEAN A_1 RESPONSE PROBABILITIES FOR LAST TWO BLOCKS OF THREE TEST TRIALS

Blocks of 3 Trials	Groups					
	C1	F1	C.5	F.5	CN.5	FN.5
4	.70	.67	.67	.70	.58	.68
5	.67	.72	.67	.68	.62	.63

comparison was made of theoretical and empirical curves for these groups.

Group differences.—Analyses of variance were performed for comparisons among groups during the varying π procedure. For each of these analyses the number of A_1 responses for each *S* was computed for each of three blocks of five test trials. This grouping of responses permitted a "between-within" type of analysis.

A comparison of Groups C1 and F1 showed significant between-group differences ($F = 11.77$, $df = 1/38$). For groups having zero probability of A_1 ($pA_1 = 0$) at the beginning of the varying π procedure there is a difference in acquisition of A_1 responses as a function of forced-choice and correction procedures.

A comparison of Groups C.5 and F.5 showed significant between-group differences ($F = 7.48$, $df = 1/38$). This indicates that for groups having "pretrained" $pA_1 = .5$ there was a significant difference in acquisition of A_1 responses during varying π as a function of forced-choice and correction procedures. Similarly, for Groups CN.5 and FN.5, which were not given pretraining but were assumed to have initial $pA_1 = .5$ at the beginning of the varying π procedure, there was a significant difference between groups ($F = 5.47$, $df = 1/38$).

Comparisons of pretraining and no pretraining procedures showed no significant difference between the two "correction" groups ($F = 3.46$, $df = 1/38$), or between the two "forced-choice" groups ($F = 0.92$, $df = 1/38$).

Table 1 shows pA_1 for each group on the last two blocks of three test trials. Since $\pi = .67$ (approximately) for daily training trials corresponding to the two blocks of three test trials, the expected value of $pA_1 = .67$. No statistical test was available to compare expected and obtained probabilities for each group but the obtained probabilities appeared to be in good agreement with the theoretically predicted value of .67.

Table 2 shows pA_1 for each group on the last 10 test trials of the varying π procedure. The table shows the number of *S*s in each group exhibiting each of the possible values of pA_1. The table shows a tendency for some *S*s in the forced-choice groups to elicit $pA_1 = 1$ on these test trials. However, no *S*s in the correction groups show this tendency and it appears that *S*s in these groups were approximately normally distributed around π. The choice data for the correction groups on the last five days of three daily training trials was also examined for tendencies to $pA_1 = 0$ or 1, but since this data was in agreement with that reported above for test trials for correction groups it is not reported.

TABLE 2

PROBABILITY OF A_1 RESPONSES (pA_1) ON LAST 10 TEST TRIALS DURING VARYING π PROCEDURE

Group	Number of *S*s Showing pA_1										
	0.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
C1				1		1	5	12	1		
F1				1			4	6	4	4	1
C.5						1	9	7	3		
F.5						3	5	5	2		5
CN.5					2	2	8	6	2		
FN.5		1				3	3	7	3	2	1

DISCUSSION

The primary purpose of the present study was to test two predictions of Estes' statistical learning theory for the special case of changing probability of reinforcement (π) in which π approaches an asymptote exponentially; viz., the predictions concerning learning rate and terminal response probability. The equation describing these predictions has been provided elsewhere (Estes, 1957, Eq. 2C). Two distinct results were obtained during the varying π procedure: (*a*) the correction procedure groups performed in good agreement with the theoretically predicted values, and (*b*) the forced-choice procedure groups far exceeded the response probabilities predicted by the theory during the early trials of the varying π procedure.

Learning curves.—As shown in Fig. 2 the predictions concerning learning rates for the three correction groups were in close agreement with the empirically obtained results. An inference from the theoretical equation is that, during the early course of learning there should be a "lag" in response probabilities behind the corresponding π values. For example, the value of π on Trial 19 is approximately .66. Given a θ value of .067 (the value for Group C.5) the theoretical equation predicts $pA_1 = .59$ on Trial 19. A similar lag was predicted for other early trials. Since the learning rate predictions were in close agreement with the results for the three correction groups, it may be inferred that there was, in fact, a "lag" of response probabilities behind the corresponding π value.

The difference between the learning curves for the correction and forced-choice groups emphasizes the importance of the fact that it is not simply the number of reinforced trials for a particular response that is important in predicting learning rates, but also the manner in which the *S*s are permitted to sample the stimulus situation. Apparently the statistical learning model is adequate to describe learning rates when *S*s are subjected to a correction procedure. Spence's model (1956) predicts that response probability to the more frequently reinforced bar would, at maximum, exceed the probability of reinforcement before declining to approach chance. Although it makes no specific predictions for cases of variable π, or for situations where a correction procedure is employed, the Spence model appears to describe the early course of acquisition for two-choice learning when a forced-choice procedure is used.

Terminal response probabilities.—According to the theoretical equation the predicted terminal response probability for each group will be the asymptotic value of π; in this case .67. From an examination of Fig. 2 and the data given in Table 1 it may be inferred that all groups reached substantially the same terminal response probability and that this value closely approximated the asymptotic value of π. Although these data substantiate the matching prediction, the forced-choice groups shown in Fig. 2 may not have reached asymptote and it appears that further training might have produced a continued downward trend. Furthermore, the interpretation of the forced-choice test trial behavior is complicated by the fact, as shown in Table 2, that some forced-choice *S*s have tended to $pA_1 = 1$. This clustering of responses to one alternative or the other is similar to the behavior obtained when a "noncorrection" procedure is employed (Hickson & Carterette, 1955; Meyer, 1960; Wilson & Rollin, 1959).

Pretraining vs. no pretraining.—The third prediction was that there would be no difference in rate of learning or terminal response probability during the varying π procedure between groups having "trained" or "assumed" initial response probabilities. No differences were, in fact, found among correction groups or among forced-choice groups which had "trained" or "assumed" initial response probabilities. The implication of this finding for statistical learning theory is that, for the two-choice situation at least, the assumption of the independence of initial response probability and θ is valid in the sense that a given initial response probability plays the same role in determining the form of a subsequent learning curve regardless of whether it was established by previous training inside or by factors outside of the given experimental situation.

SUMMARY

An experiment tested the adequacy of the following predictions from Estes' statistical learning theory: (*a*) a function derived from the theory would adequately describe rate of learning and terminal response probability in a two-choice situation in which probability of reinforcement was varied according to an exponential function; (*b*) there would be no difference between "correction" and "forced-choice" groups; and (*c*) there would be no difference between groups having "trained" and "assumed" initial response probabilities.

The *S*s, 120 female albino rats, were given bar-press training, with food reinforcement, in a runway to establish stable bar-press latencies. They were then assigned to six equal groups. Four groups were given pretraining trials is a two-choice apparatus with one of two conditions of training and of probability of reinforcement; the remaining two groups were given no pretraining. Following this, all groups were given three trials per day for 30 days in the two-choice apparatus with probability of reinforcement changing over trials according to an exponential function from .50 to .67. Thus, *S*s that had received no reinforced trials to one bar during pretraining were now reinforced on from 50% to 67% of trials for that bar-press response, and 50% to 33% for the previously reinforced response; *S*s that had 50% of pretraining trials reinforced to each bar were now reinforced on 50% to 67% of trials for one bar-press response, and 50% to 33% for the other bar-press response.

For the training series with varying probability of reinforcement, the predictions concerning learning rates for the three correction groups were in close agreement with the obtained results. The behavior of the three forced-choice groups did not confirm theoretical predictions concerning the slope of the acquisition curve. However, all groups appeared to substantiate the prediction that asymptotic response probability should match asymptotic reinforcement probability. No differences were found between groups having "trained" and "assumed" initial response probabilities.

REFERENCES

BRUNSWIK, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, 25, 175–197.

ESTES, W. K. Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika*, 1957, 22, 113–132.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill. 1959.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HICKSON, R. H., & CARTERETTE, T. Asymptotic response probability under two conditions of random reinforcement, using a noncorrection procedure. Paper read at Midwestern Psychological Association, Chicago, April 1955.

LAUER, D. W., & ESTES, W. K. Observed and predicted terminal distributions of response probability under two conditions of random reinforcement. *Amer. Psychologist*, 1954, 9, 413. (Abstract)

MEYER, D. R. The effects of differential probabilities of reinforcement on discrimination learning by monkeys. *J. comp. physiol. Psychol.*, 1960, **53**, 173–175.

RAMOND, C. K. Performance in selective learning as a function of hunger. *J. exp. Psychol.*, 1954, **48**, 265.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WILSON, W. A., JR., & ROLLIN, A. R. Two-choice behavior of rhesus monkeys in a noncontingent situation. *J. exp. Psychol.*, 1959, **58**, 174.

(Received May 20, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 145–152

VERIDICAL PERCEPTIONS OF CYLINDRICALITY: A PROBLEM OF DEPTH DISCRIMINATION AND OBJECT IDENTIFICATION[1]

PATRICIA CAIN SMITH AND OLIN W. SMITH

Cornell University

The experimental literature reads as if the topics of perceived form and perceived depth were independent and unrelated. This is true in one respect since the research on perceived forms is principally the study of plane figures while the research on perceived depth includes a host of problems related to the depth experience. In everyday life, however, man perceives the familiar objects which constitute his everyday environment, and these perceptions involve depth and form concomitantly. The two aspects of perception may be reported separately by an *O*, but in many cases perceived depths are necessary for veridical perceptions of the shapes of objects. For instance, a sphere that is illuminated only on the side away from *O* is perceived as a flat circular disk. When fully illuminated, a sphere is perceived; perception is veridical. The demonstration shows that discriminable differences in depth on a surface are necessary for veridical perceptions of at least one class of objects.

The purposes of the observations to be reported here were: (*a*) to determine whether some of the stimuli for perceived depths over surfaces were sufficient for veridical perceptions of the full convexity of a cylindrical surface in the absence of views of the lateral edges of the test object, a semicylinder (a semicylinder was used as the test object since stimuli for depth are difficult to control on spherical surfaces while easily varied on cylindrical surfaces); (*b*) to compare the relative potency of surface stimuli for depths alone with their potency in conjunction with views of the edges of the cylinder as stimuli for veridical perceptions of the shape of the object.

The veridical judgment of three-dimensional shape was set as the criterion since we assume that man perceives veridically the three-dimensional shapes of the familiar objects of his everyday environment. The sufficiency of various stimuli for mere impressions of depth has been adequately demonstrated. Reports of degrees of curvature less than those for full cylindricality are therefore of little theoretical or practical importance; the determination of conditions for veridical judgments of shape is more interesting and relevant to problems of perception of daily life.

Two different modes for designating stimuli for depth are currently in vogue. The first mode is composed of the terms for the traditional monocular and binocular cues for depth (for a recent listing see Dember, 1960). The second mode was proposed by Gibson (1950) and consists of gradients of optical texture. No suitable taxonomy exists for classifying the different optical textures which may serve as stimuli for impressions of depth and judgments of shape. The categories of the traditional taxonomy will therefore be used for naming the stimuli which were investigated. The following "cues" were investigated: linear perspective, interposition, light and shade, texture, form transformations, monocular movement parallax, and binocular disparity. Distance of observation was varied. The edges of the test object, a horizontal semicylinder, could be obscured or exposed to view, thus permitting control of its lateral contours independently of the stimuli for depth on the surface of the cylinder.

---

[1] Our thanks are due to A. E. Michotte for sponsoring and providing space for this research, and to members of the laboratory of The University of Louvain for their countless observations. The study was supported in part by: The Psychological Laboratory of Louvain, the Belgian American Educational Foundation, and Contract Nonr-401(14) between Cornell University and the Office of Naval Research.

METHOD

The apparatus has been described previously (Smith & Smith, 1957). In brief, a semicylinder (one half of a kymograph drum, radius 7.96 cm., length 16.2 cm.) was mounted with its long axis horizontal. The *O*'s line of regard (right eye for monocular, and cyclopean eye for binocular, observation) intersected the curved surface at the coincidence of the horizontal and perpendicular bisectors of the surface. Baffles, adjustable for either monocular or binocular observation, could be adjusted so as to show or to hide the lateral edges of the cylinder, but included a narrow strip above and below the cylinder behind which a uniformly gray ground was seen. Patterns (inked or three-dimensional) were placed on white paper jackets which were fitted to the 180° exterior of the cylinder facing *O*. A plain, white, unmarked jacket reflected to *O*'s eye(s) light of equal intensity (0.12 ft-L) from all portions of its visible surface except when otherwise noted. All observation excluded views of the lateral edges of the cylinder except as noted. The cylinder could be rotated on its horizontal axis. When static or in motion and covered with the plain white jacket the surface was judged to be flat by all *O*s. When in motion a flat surface appeared to expose or wipe out a gray field in the same plane as the surface. This demonstration proved that whatever curvature was reported was due either to the stimuli for depth placed on the jackets or to views of the lateral edges of the cylinder in conjunction with surface stimuli for depth.

The *O*s were seven members of the laboratory at the University of Louvain. Each was instructed to report verbally what he saw in the test area as well as its shape. When *O*'s reports required clarification, he drew pictures of what he saw. For any one test pattern *O* could observe as long as he wished. His head was correctly positioned before each view, and a shutter controlled by *E* permitted observation only when the head was correctly positioned. Unless otherwise noted all observations were binocular and were made at 116.3 cm. from the front center of the surface of the cylinder. All *O*s had periodic unrestricted views of the cylinder so as to reinforce their standard for judgment of cylindricality.

RESULTS

Linear Perspective

Linear perspective (LP) is commonly called a cue for depth. Its effectiveness as a cue has been known for centuries as demonstrated by the excellent reversible perspectives found in the Temple of Apollo at Pompeii, but little is known about LP as a stimulus for perceptions of curvature. The stimulus was investigated by varying the location, number, and width of lines placed on jackets so as to appear to be encompassing the cylinder from top to bottom when the jackets were affixed to the semicylinder.

Jacket 1.—A line 0.5 mm. wide encircled the circumference of the semicylinder at its horizontal midpoint. When observation was monocular, the surface was reported by all to appear to be completely flat. With binocular vision slight curvature of the surface was reported by some *O*s, but these judgments were consistent neither within nor between *O*s.

Jacket 2.—The width of the center line was gradually increased to 2.5 cm. for both monocular and binocular views. Some curvature was reported by all *O*s with monocular views, but a greater degree of curvature was reported by all *O*s with binocular views. No judgments were obtained of full cylindricality. At best, the center front horizontal portion of the area seemed to be flattened, while the top and bottom limits of the cylinder appeared to extend parallel to the line of sight.

Jackets 3 and 4.—Increases in the width of the center line in steps of 0.5 cm. beyond 2.5 cm. resulted in judgments of less curvature by all *O*s. The optimal visual angles to be subtended by a solid line for the discrimination of the curvature of a cylinder remain to be investigated. These visual angles, when determined, will probably vary with the radius of the cylinder.

Jackets 5–12.—Inked bar lines were centered on the cylinder. The widths of these bar lines varied in steps of 0.5 cm. from 1 cm. to 2.5 cm. For each bar line there was a pair of narrow (0.5 mm. in width) lines which enclosed the same area and were correspondingly positioned on the cylinder on separate jackets. Judgments of curvature induced by the bar lines were compared with judgments of curvature induced by corresponding pairs of parallel lines. In all comparisons the solid bar lines were more effective and yielded reports of greater curvature.

Jacket 13.—Two vertical 1-mm. lines were placed so as to obtain the maximal possible horizontal separation on the cylinder. They were judged to be curved and to "toe in" at either end although the surface on which they were seen was reported flat. This demonstrates that parallel lines which recede in depth from *O* do not necessarily induce perceptions of depth. Geometrically, perspective was present, but in the absence of stimuli sufficient for perceived curvature of the surface, the lines were judged as curved, a response appropriate for the proximal stimuli, but not for the distal stimuli. This class of result is not uncommon in studies of size constancy.

Jackets 14–17.—Two, 13, 27, and 54 vertical 1-mm. lines were evenly distributed over the horizontal extents of separate jackets. No veridical judgments were reported. Even with the 54-line jacket the horizontal middle of the cylinder appeared to be flattened. Veridical judgments were reported by all *O*s for both the top and bottom of the cylinder where the physical changes in depth were relatively greater per unit of visual angle than in the center. From this may be concluded that LP is an effective stimulus for veridical perceptions of changes in depth over surfaces where depth differences in terms of visual angles are fairly rapid.

Texture

The vertical lines minimized the effects of texture in the optic array. Lines in the vertical dimension were therefore replaced by columns of dots so as to maximize texture and its packing density in the optic array as a function of receding surfaces from *O*.

Jacket 18.—Five round dots 0.5 mm. in diameter with equal vertical separations were placed in line on the center vertical of the cylinder. The *O*s reported the dots as aligned, but without the phenomenal continuity characteristic of a line. No curvature was reported.

Jacket 19.—Seven similar dots distributed as above yielded the same results.

Jacket 20.—The diameters of the dots of Jacket 19 were increased to 5 mm. The spots appeared to be a line to all *O*s and some curvature of the surface was reported.

Jackets 21–23.—The diameter of the dots was held constant at 5 mm. while two columns were presented in which the number in each was increased until 25 were aligned vertically and parallel to a second line of 25 dots. Two centimeters separated the interior edges of the two lines which were centered on the cylinder. Two parallel lines of dots were only slightly more effective than a solid line, 3 cm. wide. Veridical judgments were not obtained. Monocular views always resulted in less apparent curvature than did binocular observation. In this case the combination of binocular disparity, LP, accommodation, convergence, and texture was not sufficient for veridical judgments of the shape of the cylinder. These results indicate that the gradient theory of optical texture may be geometrically exact, but psychologically insufficient for accounting for veridical perceptions of shape in all situations. Optical texture and its gradients are sufficient under some conditions for perceptions of depth and may therefore be called cues in the same sense that parallel lines which induce impressions of depth are designated as the cue of LP.

Vertical parallel lines were sufficient for veridical judgments of the shape of the test surface at the top and bottom of the cylinder even though the center of the space appeared to be flattened. The consequent hypothesis was that curvature reported in any one part of the cylinder is independent of curvature reported for other parts.

Jacket 24.—The above hypothesis was tested by placing 11 vertical lines, 1 mm. wide, 3.0 cm. long with horizontal separations of 1.5 cm., at both the top and bottom of the cylinder. Each line at the bottom was in line with a corresponding line at the top. Compelling impressions of curvature were induced at both extremes for all *O*s, but the intermediate area joining the apparently curved areas was reported as flat by all.

Jackets 25–29.—The same lines at both the top and bottom were progressively lengthened on separate jackets until only a narrow, horizontal strip in the center of the cylinder, 3.0 cm. high, prevented closure of the lines. This narrow strip was judged to be completely flat by all *O*s. The absence of stimulation for depth resulted in reports of apparent flatness of the surface, and this in turn caused an apparent change of direction of the surface to the frontal parallel.

Koffka (1935, p. 153) states that the gestalt law of good continuation is applicable to curved lines. This law is inapplicable to the curved lines defining curved surfaces in perception. The law, if functional, may be appropriate to curved lines on plane surfaces where the curvature of a line may be identified independently of the form of the surface on which it exists.

Density of Texture

Gradients of texture in the optical array are an invariant function of perspective geometry when surfaces receding in distance from *O* are viewed by *O*. Since the gradients always exist they are potentially descriptions of light in the optic array as reflected from diverse collections of objects such as leaves, people, trees, and sandy stretches. Gibson (1950) proposed optical texture as an effective stimulus for recession of a surface (apparent depth). In addition to being a cue for apparent depth, are optical textures also sufficient for veridical judgments of cylindricality? The displays to be described resulted in samples of some but not all optical textures.

Jacket 30.—Horizontal and parallel lines, 0.5 cm. wide and separated vertically by 2 cm. covered the cylinder. This type of optical texture minimized the effects of binocular disparity. Judgments varied between and within *O*s. Some would report some curvature, and then report stairsteps on the ascending side (top) of the cylinder. The perceptions were unstable except in the case of one *O*, who consistently reported "fairy stairsteps."

Jacket 31.—The number of equally spaced horizontal lines was doubled. Some curvature was reported by all, but the center of the cylinder appeared to be flat. The staircase effect was still obtained, but less frequently. No veridical reports of cylindricality were obtained.

Jacket 32.—Horizontal rows of dots, 0.5 mm. in diameter and separated vertically by 1 cm. were presented so as to increase the effects of binocular disparity. The dots were haphazardly spaced on the horizontal so as to minimize the appearance of vertical alignment. The staircase effect was not obtained, and all *O*s reported some curvature, but no veridical judgments. The cylinder also appeared to be flattened in the middle. More curvature was reported with the dotted lines than when they were solid, demonstrating effects of binocular disparity.

Jacket 33.—LP was enhanced by creating columns and rows of 1-mm. dots separated by 1 cm. Some, but incomplete, curvature was again reported.

Jackets 34–39.—The position and number of dots were held constant while their diameter was increased to 5 mm. This resulted in an increase in apparent curvature for all *O*s. When the dots were enlarged to about 7 mm. in diameter, however, the apparent curvature had seemingly reached a maximum, and a further increase in spot diameter to 1 cm. resulted in less reported curvature. The size of the dots was then held constant at 5 mm. in diameter while their density was varied. On each of a series of jackets the dots were uniformly distributed, but with no apparent order with respect to vertical or horizontal alignments. Dots separated from adjacent dots by an average of 1 cm. always resulted in the appearance of more curvature than did dots separated by only 1 mm.

Several factors seemed to be functioning in determining the degree of reported curvature or its equivalent, the amount of depth that was judged. The size and number of dots per unit of visual angle seemed to be interacting with the physical extent in depth of the surface per unit of visual angle. The interactions which result in optimal judgments need to be determined empirically. In the above series of observations different optical textures, binocular disparity, LP, accommodation, and convergence in combination were not sufficient for veridical judgments of the shape of the cylinder.

Form Transformations

The form transformations of both regular and haphazardly arranged circular dots were ineffective for veridical judgments as described above. It was hypothesized that the transformations of rectangles to trapezoids might be more effective.

Jacket 40.—Rectangles 1 cm. wide and 2 cm. long were arrayed in vertical columns lengthwise. The columns were separated horizontally by 1 cm., and the width of lines was 0.5 mm. Clear impressions of curvature were reported, but in all cases these did not approximate the true shape of the surface.

Interposition

Jackets 41–42.—Irregular cardboard inserts were placed in slits in the jackets so that the plane of each insert was perpendicular to the surface of the cylinder. Both high and low frequencies of interposed forms were used. Interposition was sufficient for valid judgments of relative depth, that is, one insert could be judged as being in front of another, but little curvature of the form itself was reported.

Light and Shade

Gross differences in brightness of light reflected to the eye from the jackets were introduced for all jackets so far described. Differences in brightness on the surface of plain, white, unmarked jackets did not result in judgments of curvature. Differences in brightness in conjunction with patterned surfaces had, in general, the effect of increasing the degree of reported curvature, but these effects were not uniform. The pattern composed of rectangles in columns seemed to be most affected. No veridical judgments resulted from the introduction of this cue in conjunction with the previously reported cases.

Monocular Motion Parallax

This cue was tested in conjunction with others by rotating all previously described patterns during monocular observation. No reports of the true shape of the cylinder were obtained, and substantial increases in perceived curvature were not reported. This method of generating monocular motion parallax is not typical of the manner in which it is usually caused in everyday life, since it most commonly results from motions of the head. Consequently, these results should not be considered as typical of those which might be obtained from normal motions of the head.

Combined Factors

Jacket 43.—A repetitive pattern of 14 columns of rectangles, each enclosed by two vertical parallel lines was tested. All lines were 1 mm. wide. The rectangles were $\frac{1}{2}$ cm. wide by 1 cm. long. A vertical line 3 mm. long connected the center of the top of each rectangle to the center of the bottom of the rectangle above it. The parallel lines were separated by $1\frac{1}{2}$ mm., and the long side of each rectangle was 2 mm. from the near side of one of the parallel lines. Good contrasts in brightness (light and shade) existed on the jacket. Veridical judgments were obtained from all *O*s. In this situation there were good LP, form transformations, binocular disparity, light and shade, texture, accommodation, convergence, but no interposition or motion parallax. The distance of observation was then increased to 232.6 cm., and the surface no longer appeared to be cylindrical, but still appeared to be curved. The distance of observation was then increased to 348.9 cm., when the surface appeared to be flat to all *O*s.

Other Veridical Reports

One *O* who had had monocular vision since early childhood visited the laboratory and was tested on the textures described in the section on LP. She spontaneously and correctly judged the shape of the surface when 5-mm. dots were separated from adjacent dots by as much as 1.5 cm. This suggests that: (*a*) an important source of individual differences in perception is the state of learning of *O*, and (*b*) it is possible that any combination of factors may be sufficient as stimuli for veridical perceptions when sufficient learning has occurred.

Knowledge of Shape of Cylinder

Knowing that the surface to which the jackets were affixed was actually semicylindrical had no detectable effect on the appearance of the test area. For each series, the reports of naive *O*s were checked against those made by the 7 *O*s of this study. No appreciable differences in reports were obtained. These comparisons also served as a practical control for order effects in the large number of observations.

View of Lateral Edges of Cylinder

Each of the jackets was displayed with a wider aperture, so that the lateral edges of the cylinder were visible. All of the jackets except 1 and 18 were readily discriminable as cylindrical. The full contour was not sufficient for veridical judgments when the cylinder was covered by a plain white jacket. Conversely, in the absence of views of the lateral contours the cues for depth were in most instances insufficient for veridical reports. Surface cues for depth in combination with the lateral and vertical outlines of the cylinder were, however, consistently effective for veridical reports.

DISCUSSION

The lateral edges of the cylinder in conjunction with the surface cues for depth were highly effective in inducing veridical perceptions of cylindricality. The edges of the cylinder were not sufficient for veridical judgments of its shape in the absence of other stimuli for depth. They were, however, remarkably potent stimuli for the veridical judgment in conjunction with stimuli for depth which were not sufficient independently of the edges.

The lateral and horizontal edges clearly delineated the cylinder as an object in perception with its own edges. This is normally the case in everyday life where depth is seldom perceived independently of three-dimensional objects. In the laboratory, however, the custom of the single variable experiment has dictated that the edges of a test space be defined by the edges of apertures. Judgments of slant, tilt, rotation, and curvature under aperture conditions may represent only generalizations by *O* from complete to impoverished stimulation, with the consequence that errors of judgments should be expected to be in excess of those normally expected in everyday life. If this is the case, a demonstration that variation in stimulation is accompanied by correlated responses can result in a quantitative science of effects, but these effects will serve as only qualitative predictors of responses in everyday life. For instance, given specified gradients of optical texture, the prediction should be valid that depth will or will not be perceived, but knowledge of the effect on responses of such gradients in the laboratory under aperture conditions should not be expected to predict the magnitude of errors of responses to equivalent gradients encountered in everyday life.

In the absence of stimuli for depth, *O*s reported merely flat rectangles extending across the width of the test area. This confirmed a prediction by Hochberg and Smith (1955) that other things being equal, the depth judgment requiring the least perceptual information will be the one that is made. That flat rectangles were perceived in the absence of cues for depth is evidence that less perceptual information is required for perceiving a flat surface than for perceiving a curved surface.

The literature on accommodation, convergence, motion parallax, and binocular disparity as cues for depth is substantial, but this literature is rarely relevant to the perceived depths required for veridical perceptions of the shape, slant, or tilt of three-dimensional surfaces. All other things equal, common sense dictates that binocular views of textured surfaces should result in better judgments than monocular views, and this has been confirmed by Clark, Smith, and Rabe (1956) and by Smith and Smith (1957). However, that which is known about the above cues permits neither the prediction of how much better judgments should be with binocular than with monocular views, nor specification of the circumstances in which binocular judgments should be better than monocular ones.

The infrequency of veridical judgments of shape of the cylinder when lateral contours were occluded might be attributed to aperture effects which may cause a surface to appear to come up into the plane of the aperture (Katz, 1935). The aperture of this study permitted a vertical field of view of 7° 20′ when *O* was positioned 116.3 cm. from the front surface of the cylinder. This field was larger than the 7° field of Gruber and Clark (1956) who found 7° sufficient for minimizing the "come-uppance" phenomenon. Aperture effects were therefore not considered to be important in our results.

The report of the monocular *O* indicates the possibility that the patterned stimuli sufficient for reports of depth as well as veridical judgments are those patterns which have been learned by *O*s as signifying cylindricality. Should this be the case any of the stimuli is potentially effective for eliciting veridical judgments. Effectiveness would depend upon experience with the particular stimulus dimension in combination with other information concerning cylindricality.

SUMMARY

Some of the stimuli for perceived depths on surfaces were tested for their effectiveness in inducing veridical judgments of cylindricality. Views of lateral edges of the test object were also tested independently and in combination with stimuli for depth. The stimuli of texture, form transformation, linear perspective, accommodation, convergence, light and shade, and binocular disparity in

combination were sufficient for veridical perceptions of the shape of the cylinder. The lateral (and horizontal) edges of the cylinder alone were not sufficient, but in combination with stimuli for depth they were highly effective in inducing veridical perceptions of the shape of the object. One *O* who had had monocular vision since early childhood reported the shape of the test object correctly in situations in which normally sighted *O*s with binocular views reported lesser degrees of curvature. Hence, any of the stimuli for depth may be potentially effective for veridical judgments depending on the status of learning of *O*.

REFERENCES

CLARK, W. C., SMITH, A. H., & RABE, A. Retinal gradients of outline distortion and binocularity as stimuli for slant. *Canad. J. Psychol.*, 1956, **10**, 77–81.

DEMBER, W. N. *Psychology of perception.* New York: Holt, 1960.

GIBSON, J. J. *Perception of the visual world.* New York: Houghton Mifflin, 1950.

GRUBER, H. E., & CLARK, W. E. Perception of slanted surfaces. *Percept. mot. Skills*, 1956, **5**, 81–92.

HOCHBERG, J., & SMITH, O. W. Landing strip markings and the "expansion pattern": I. Program, preliminary analysis, and apparatus. *Percept. mot. Skills*, 1955, **5**, 81–92.

KATZ, D. *The world of colour.* London: Kegan Paul, 1935.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

SMITH, O. W., & SMITH, P. C. Interaction of the effects of cues involved in judgments of curvature. *Amer. J. Psychol.*, 1957, **70**, 361–375.

(Received May 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 153–157

COPYING REDUNDANT MESSAGES [1]

JERRY A. HOGAN

Harvard University

From studies of the recall of non-random sequences of English words (Miller & Selfridge, 1950), of nonsense syllables (Aborn & Rubenstein, 1952; Rubenstein & Aborn, 1954), and of letters of the alphabet (Miller, 1958), it is quite clear that the rules of formation of these sequences greatly influence both the amount of material and the amount of information which can be recovered after a given presentation of this material. This and related work has been summarized by Rubenstein and Aborn (1960). Several of the above studies were designed to test the hypothesis that the amount of information proccessed and retained is invariant under changes in the rules of formation; that is to say, that the redundancy of the material does not affect the amount of information processed and remembered. This hypothesis was not supported. The relation between redundancy and the amount of information or the amount of material processed has not been determined.

In looking for such a relation, it is reasonable to use sequences generated by finite state grammars (see Chomsky & Miller, 1958; Miller, 1958) since both the redundancy of these sequences and *S*'s pre-experimental experience can be stated precisely. Even in these sequences, however, the number of parameters involved, such as number of alternatives (letters), number of states, rules of transition, *S*'s experience, and so on, makes a complete exploration of redundancy difficult. A start can be made by using a relatively simple grammar with a small number of alternatives. The relation that exists between the redundancy of the messages and the amount of information processed could then be used as a baseline with which to compare the results using other parameters.

The present study was designed to investigate this relation. However, instead of using one of the familiar measures of retention to indicate how much of the input information had been processed, a copying task similar to that used by Sumby and Pollack (1954) was used. In their study *S*s copied messages by glancing at different word-order approximations to English, turning away, and writing down the material they remembered. In the present study *S*s processed information by copying messages written in artificial languages of varying redundancies. Two experiments were performed. In Exp. I naive *S*s were used to determine the effect of redundancy on their copying behavior. In Exp. II *S*s had repeated opportunities to become familiar with the rules of formation of the various languages. The copying behavior of these experienced *S*s was compared with that of the naive *S*s.

METHOD

Experiment I.—A finite state grammar with varying transitional probabilities (see Fig. 1) was used to generate strings of symbols of four redundancies: 53.3%, 75.9%, 83.6%, 89.6%. Essentially random strings (3.5% redundant) were constructed with the

---

[1] Part of this work was carried out while the author was under the tenure of a National Science Foundation Predoctoral Fellowship. The author wishes to thank George A. Miller for his helpful suggestions and advice.

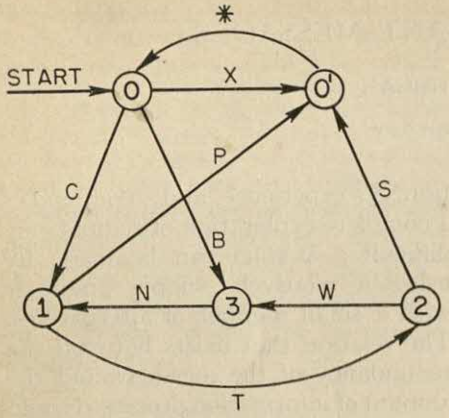


FIG. 1. Diagram of finite state generator of the 75.9% redundant messages used in these experiments. (In each state, all the transitions to each possible next state are equally probable. These transitional probabilities were altered—and the symbols generated in transition were switched—in constructing the more highly redundant messages. The 53.3% redundant messages were generated by doubling the choice of symbols in each transition: e.g., the transition from State 1 to State 2 could generate either T or C, etc.)

use of a random number table. The vocabulary consisted of nine symbols: eight consonants and an asterisk. The set of strings of each redundancy constituted a language. A sample from each language is given in Table 1. Four messages of 134 characters each were constructed from each language.

All the messages were typed on 5 × 8 in. cards. Each card contained two messages from the same language.

The apparatus was a box, the front of which was a one-way glass. When *S* pushed a button, a light came on inside the box and the material on a single card was visible. A counter registered the number of pushes *S* made, and a Springfield timer registered the number of seconds the button was held down.

The task of each *S* was to copy the messages in front of him. He wrote letters down only when not looking at the message. His instructions were to "be as accurate as possible . . . but do not spend time checking for errors. Speed is not important, but try to work at a steady pace."

Ten college students were used as *S*s. All were completely naive about the experimental situation: they had had no previous experience with the kinds of languages used. Each *S* copied all 20 messages. A balanced latin square was used as the design. After each *S* had copied 10 messages he was given a 5-min. rest. The total time required was about 1½ hr. per *S*.

Experiment II.—The *E* and two friends were used as *S*s. The *E* had constructed the messages, carried them through their various stages—proofreading and typing—and had checked some of the copied messages for mistakes. With this background, he copied 10 of the 20 messages used in Exp. I.

The other two *S*s were originally naive, and their performance in copying messages was studied over a period of time. After copying one message of each redundancy, the principles underlying the construction of the languages were explained to them. They were also shown the explicit patterns which they would find in each language. Following this instruction, over a period of two months, each devoted 11 hr. to copying messages and to checking each other's papers for mistakes. During each 1 hr. experimental session, they usually copied and checked two messages from each language. The 20 messages used in Exp. I were normally used. However, new messages from these same languages were used as Messages 13, 14, 22, and 23 from each language. Each *S* copied a total of 23 messages from each language (115 messages in all).

RESULTS

Experiment I.—For naive *S*s there is evidence for a small amount of learning from the beginning to the end of the session, probably associated

TABLE 1

SAMPLES FROM LANGUAGES OF VARIOUS REDUNDANCIES

3.5%	BS*BX*CNW*X*CPB*TNX*B*CBBXXPC* WCXT*NBCCN*S*NWC*BN*XW*CBXPCT*
53.3%	N*N*CP*SP*CCNXP*N*BXTNNP*BNT*W NCWXT*BXP*SCWXT*X*X*N*WXT*X*N*
75.9%	X*X*X*BNP*X*CTS*BNP*X*BNTS*CTW NP*X*BNP*BNTS*CTWNTWNTWNTSNP*B
83.6%	N*WXT*SCBXCBXT*ST*N*N*WXCP*WBC BXCBXCBXT*WXCBXCBXCBXCBXCBX
89.6%	NPBSPBSPBSPBSPBSPBSPBSPBSPBSPB SPBSW*NPBSPW*C*NPBSSBPPBSPBSPB

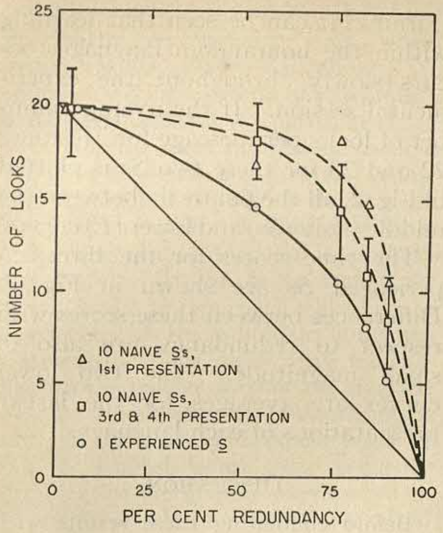


FIG. 2. The number of looks required to copy 134-symbol messages of varying redundancy. (The upper curve has been multiplied by a constant factor of 0.96; the lower curve has been multiplied by a constant factor of 1.16. The variability of the values for the naive *S*s on the third and fourth presentations has been indicated by adding ±σ to each point.)

with familiarity with the task and the general situation.

There also seems to be some learning within each nonrandom language through the four presentations. This can be seen from the two upper curves of Fig. 2. The top dashed curve gives the average number of looks for the first presentation of each language. (This curve has been corrected for learning associated with the situation by multiplying each value by a constant to bring the points at 3.5% redundancy into coincidence.) It can be seen that three of the four remaining points are considerably above the corresponding points on the second dashed curve. This latter curve gives the average number of looks for the third and fourth presentations of each language. (Using a *t* test on the corrected values, $P < .01$ that the points for languages of 76% and 84% redundancy are from the same distributions.)

There are striking differences among the various languages with respect to the average number of looks necessary to transcribe messages of equal length. The number of symbols carried per look increases substantially with the increase in redundancy. In terms of a measure of information, however, the amount carried per look decreases with an increase in redundancy since the number of symbols carried per look does not increase as rapidly as the information per symbol decreases.

A sample of errors indicates that they are relatively infrequent: about one per message, and less in the 89.6% language. These errors were not taken into account when values of information were calculated.

The time scores are given in Fig. 3. The dashed curve gives the average number of seconds that *S* looked at messages from each language on the third and fourth presentations. The differences among the languages, though perhaps statistically reliable, are not of large magnitude. Hence, since the total information is less

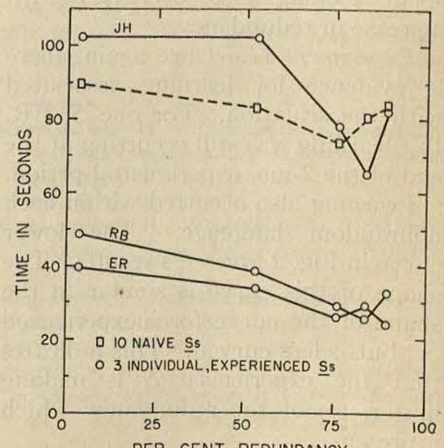


FIG. 3. The average number of seconds that *S*s looked at 134-symbol messages of varying redundancy while copying.

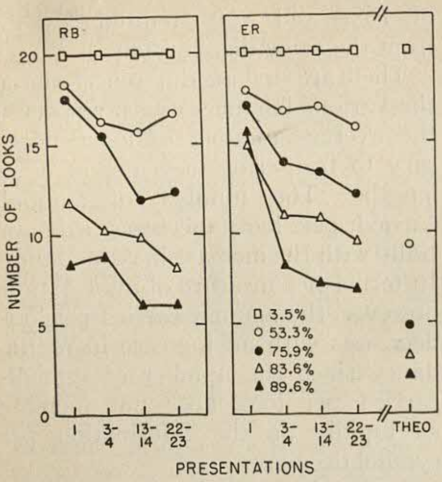


FIG. 4. The number of looks required to copy 134-symbol messages of varying redundancy as a function of the number of presentations, for 2 *S*s. (The values for each presentation have been multiplied by that constant factor which made the corrected value equal to 20 for the 3.5% redundant message. On Presentations 1, 3–4, 13–14, and 22–23, these factors were 0.83, 0.75, 0.85, and 0.85 for RB and 0.74, 0.82, 1.03, and 1.05 for ER, respectively. The values on the right indicate the theoretical asymptotes which would hold if the constant information hypothesis were true.)

with greater redundancy, the amount of information carried per unit time spent looking also decreases with increase in redundancy.

Experiment II.—Once again, there is evidence for learning associated with the situation. For one *S*, ER, this learning was still occurring at the end of the 2-mo. experimental period.

Learning also occurred within each nonrandom language. The lower curve in Fig. 2 gives *E*'s results. The shape of this curve is similar to the shape of the curve for inexperienced *S*s, but is less curved. This indicates that the experienced *S* is making better use of the redundancy which is present.

Figure 4 gives the course of this learning for two *S*s. When all the values have been corrected for overall learning, it can be seen that learning within the nonrandom languages occurs slowly throughout the experimental session. If the average number of looks per message for Messages 22 and 23 for these two *S*s is plotted in Fig. 2, all the points lie between the middle (naive *S*s) and lower (*E*) curves.

The time scores for the three experienced *S*s are shown in Fig. 3. Differences between these scores with respect to redundancy are also of small magnitude. The two lower curves are averages for the last 6 presentations of each language.

DISCUSSION

Before comparing these results with those of earlier studies, it should be noted that in most previous studies *S* had to memorize a passage and then recall it. In the present study, *S* was required to copy a passage which was visible to him only when he was not writing. Both methods can be used to study how information processing varies with redundancy, but the second method has the advantage of experimental efficiency. Using the second method, we are directly studying the effects of redundancy on immediate memory span. Whether or not these effects correlate with the effects of redundancy on longer term memory would seem to be, primarily, an experimental question. The results of this study indicate that those effects do, indeed, correlate very highly.

As in the previous studies, it was found that with highly redundant languages *S*s process more symbols but less information than with random languages. The relation of both symbols per look and information per look to redundancy can be plotted easily from the data in Fig. 2 and from the fact that the messages consisted of 134 symbols each. Using the values on the second dashed curve, information per look drops almost linearly from about 21.3 bits per look with a random message to about 4.7 bits with a message which is 89.6% redundant. The number of symbols per look starts at about seven for the random language—which is what the many experiments on memory span would predict (Miller, 1956)—and above 75% redundancy it increases rapidly to infinity as 100% redundancy is approached.

Perhaps the most interesting result is the effect of experience with the redundant languages. In Fig. 2 the straight line gives the expected number of looks if the constant information hypothesis were correct. It can be seen that the curves begin to approach this line as experience with the redundant languages is increased. This result may indicate that with extensive further training the constant information hypothesis might hold. If so, we would expect it to be true for English since these sequence dependencies are greatly overlearned in the normal *S*. If the method suggested by Shannon (1951) is used to estimate the redundancies of various word-order approximations to English, the two studies which used English words as stimuli (Miller & Selfridge, 1950; Sumby & Pollack, 1954) tend to support this hypothesis. In any case, experience certainly seems to allow greater use of the available redundancy. The data in Fig. 4 indicate, however, that this learning occurs very slowly.

The time needed to transmit a given number of symbols appears to remain fairly constant as a function of redundancy, regardless of experience. This is not surprising since *S*s normally pronounced the letters subvocally; it takes about the same amount of time to pronounce a letter regardless of context. There was another factor involved in copying messages from the two most highly redundant languages; namely, it was more difficult for *S* to keep his place because of the similarities within the message. Practice in copying seems to lower the absolute values of the time scores.

SUMMARY

A relation between the redundancy of messages generated by a finite state grammar and a measure of information processed was determined. Ten naive *S*s copied messages of varying redundancies. Both the amount of time *S*s looked at the messages and the number of times needed to look at the message to copy it were plotted as a function of redundancy. The amount of time spent looking remained fairly constant as a function of redundancy, but the number of looks decreased monotonically, slowly at first, and rapidly above 75% redundancy. If a constant amount of information had been carried per look, this curve would have been linear. These results correlate highly with the results of earlier studies which found that a constant amount of information was not recalled after a given presentation of variously redundant material.

The results of three experienced *S*s agree generally with the other results except that the relation between looks and redundancy is less curved than the same relation for the naive *S*s. This indicates that experience with redundant materials allows greater use to be made of the redundancy which is present.

REFERENCES

ABORN, M., & RUBENSTEIN, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.

CHOMSKY, N., & MILLER, G. A. Finite state languages. *Inform. Control*, 1958, **1**, 91–112.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MILLER, G. A. Free recall of redundant strings of letters. *J. exp. Psychol.*, 1958, **56**, 485–491.

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.

RUBENSTEIN, H., & ABORN, M. Immediate recall as a function of degree of organization and length of study period. *J. exp. Psychol.*, 1954, **48**, 146–152.

RUBENSTEIN, H., & ABORN, M. Psycholinguistics. *Annu. Rev. Psychol.*, 1960, **11**, 291–322.

SHANNON, C. E. Prediction and entropy of printed English. *Bell Sys. tech. J.*, 1951, **30**, 50–64.

SUMBY, W. H., & POLLACK, I. Short-time processing of information. *Hum. Factors Operat. Res. Lab. Rep.*, 1954, No. TR-54-6.

(Received May 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 158–163

RESPONSE BIAS AND PERCEPTION

CHARLES D. SMOCK AND FREDERICK H. KANFER

Purdue University

One important by-product of the controversies about the role of personality variables in perception has been the clarification of some major methodological problems associated with the concept of perception. There is now general agreement, for example, that experimental arrangements must provide for experimental or statistical separation of the relative contributions of the perceptual system and response biases to overt (indicator) responses (Garner, Hake, & Eriksen, 1956; Kanfer, 1956). This issue is most clear in the case of perceptual recognition studies where the problem is one of isolating variables that affect seeing from those that affect saying (Neisser, 1954). The influence of prior training on subsequent recognition thresholds has been demonstrated to be so strong that response bias has often sufficed to explain the experimental results without recourse to the properties of the perceptual system that might contribute to recognition. However, typical procedures for determining thresholds by the ascending method of limits produces a shift in the relative contribution of sensory input and prior established response biases (Bricker & Chapanis, 1953). Consequently, the extent of influence of each of these systems to recognition responses has been difficult to isolate.

Two quite distinct methods for separating the relative effect of the perceptual and the associative response variables to recognition have been suggested. One approach (Blackwell, 1952) involves experimental situations in which the instrumental recognition or "indicator" response varies systematically over a series of presentations of a particular stimulus; i.e., response transposition. A second method is designed to study the effects of the particular experimental procedures in the absence of perceptual input (Hochberg, 1956). Once the effects of the procedures alone are known, perceptual variables could then be systematically introduced. Goldiamond and Hawkins' (1958) recent study of the effects of pretraining on recognition responses in a quasiperceptual task is an example of this latter approach; i.e., an attempt to isolate the "pure" effects of response biases.

The present investigation was designed to determine the effects of a perceptual set on response biases in a quasiperceptual situation. Such response biases may be more or less congruent with the demands of the task situation. Additionally, there is a set to attend to external stimulation in any perceptual experiment. It might be expected that any variable which affects *S*'s expected source of available information would further influence response probabilities; these effects could not be ascribed to the "response" system or response biases.

In the present study we were concerned with two factors which would appear particularly important in restricting free utilization of previously trained responses in perceptual recognition experiments. First, instructions to "look at the screen" evoke perceptual attending responses which provide, through visual exploration, indeterminate sensory input likely to modify response biases. Second, failure to receive stimulus confirmation of identifying responses has frustrating and emotional consequences known to alter response probabilities (James & Rotter, 1958). Removal of an *E*-defined stimulus, therefore, may *minimize* the effects due to perceptual input, but does not necessarily provide a "pure" situation for the operation of previously acquired response biases.

These considerations lead to the following hypothesis. Maximum effectiveness of prior training will be manifest where: (*a*) *S*s expect no additional "external" information prior to their response; i.e., when there is no perceptual input with which the response can be brought into congruence, and (*b*) when *S*s expect minimal accuracy of recognition and thus adverse effects of negative reinforcement ("incorrect responses") are minimized. Conversely, conditions which evoke a set to perceive and which fail to provide confirmation of perceptual expectancies should restrict utilization of response biases. A third experimental condition in which the set to perceive is minimized but in which the confirmation of response accuracy is maximized, or vice versa, should result in intermediate contributions of the response biases.

In order to test this hypothesis three groups of *S*s were tested in a quasiperceptual task. The first group (ESP) was instructed that no perceptual information would be forthcoming. The second group (SLP) was told that some information would be presented and response accuracy was expected. The third group (LNG) was instructed that information must come from "internal" sources, i.e., from memory. It was predicted that relative accuracy due to response biases would be most congruent with pretraining frequency under ESP, and least congruent under SLP instructions. The LNG group was expected to fall between the other two groups.

METHOD

Subjects

The subjects were 42 male college students from an introductory psychology course.

Stimulus Materials

The stimulus words and word lists for the training series were identical to those by Goldiamond and Hawkins (1958). Ten nonsense words were arranged into five pairs which were systematically varied so that each pair was presented at five different frequencies during pretraining. Thirteen extraneous words appeared only once during training and not at all during the test phase. The paired words were: MIV, WUX; TUD, ZOF; HIF, MIF; FEP, HAJ; VUK, VOF. Five decks of 99 cards each were prepared so that word pairs and training frequencies were counterbalanced. Training frequencies used in the study were 25, 10, 5, 2, and 1. Subjects were assigned randomly to training deck and experimental group.

Training Procedures

Each word was typewritten in capital letters in the center of a 3 × 5 in. white card. The deck was shuffled separately for each *S* and placed face down between *E* and *S*. The *S*s were told that these words were from a foreign language and that *S* was to spell and pronounce the word exposed on each stimulus card. A card was presented every 8 sec.

As interpolated material, *S* was given material, excerpted from national magazines, to read during a 10-min. interval between training and the test situation. For each group, the contents of the material was directly relevant to the instructional conditions to which *S* has been assigned.[1]

The *S*s sat 6 ft. in front of a 10 × 12 in. milk glass screen centered on a 6 × 10 ft. table divider. A signal light was mounted above the screen. "Stimulus" presentation was automatically controlled. A warning bell sounded 2 sec. prior to stimulus presentation. A 7-sec. interval between stimulus

---

[1] Copies of the excerpts used here will be mailed upon request to the authors.

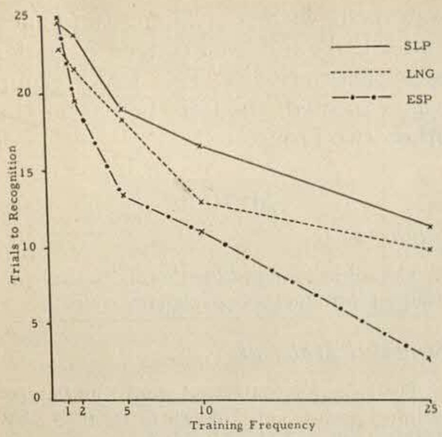


FIG. 1. Trials to "recognition" as a function of training frequency and instructions.

presentation and the bell signaling the next trial was maintained.

Instructional Conditions

Following pretraining, instructions were systematically varied in three respects: (*a*) the contents of the interpolated reading material; (*b*) the nature of the "stimulus" presented during the test phase; (*c*) the stated purpose of the experiment. The essential aspects of these variations are presented below.

Extrasensory perception group (*ESP*).—Interpolated reading material consisted of excerpts from national magazines on ESP phenomena. The *S*s were told that 2 sec. after the warning bell *E* would be thinking of one of the foreign words which *S* had pronounced earlier. When *E* was "*concentrating the hardest*" and the possibility of *S* receiving the thought wave was at its maximum, *E* would turn on the signal light. At that time *S* was to report the word image that appeared in his mind. It was further emphasized *S* should guess what word it was, even though the word appeared only dimly in his mind, since the purpose of the experiment was *to determine the extent to which* S *would receive the thought waves of another person.*[2]

Subliminal perception (*SLP*).—Instructions for this group were as similar as possible to those reported by Goldiamond and Hawkins (1958). The phenomena of subliminal perception was emphasized in introducing the reading material. The contents of the reading material was relevent to subliminal perception findings during recent years. Following the interpolated activity *S* was told that 2 sec. after the warning bell, one of the foreign words he had seen and pronounced in the training series would be flashed on the screen. The word would appear very briefly and *S* should respond, even though he felt he was guessing, since the purpose of the experiment was to relate behavior to *unseen or barely seen* stimuli.

Learning group (*LNG*).—The interpolated reading material dealt with the effects of "learning without awareness" on subsequent behavior. The *S*s were told that 2 sec. after the warning bell *E* would be ready to record his answer and would turn on a signal light to indicate this fact. The *S* was to respond with any one of the foreign words which he recalled from the earlier list. The *S* was instructed to "*search his memory*" on each presentation since the purpose of the experiment was to relate behavior to the *memory for events which* S *had not been aware of learning.*

The *S*s in all three groups were asked to respond when the "stimulus" was flashed. They were also told that only pretrained words would be used and *E* would inform them when they had responded correctly. "Recognition" in each case was defined as congruence of *S*'s emitted response with the predetermined randomly selected word which had to be guessed on each series. After a correct response or 25 incorrect trials a new series was begun. A total of 10 series, representing two words at each frequency, was administered to all *S*s.

Recall Conditions

Following the quasiperceptual task, *S*s were asked to write down all the words they could remember from the pretraining series.

RESULTS

Trials to recognition.—Trials to recognition represents the number of responses made before *S* emitted a response that matched the one selected by *E* for that series. The relationship between pretraining frequency and trials to recognition for the three experimental groups is presented in Fig. 1.

In order to reduce the skewness of

---

[2] For the information of the purists, *E* did concentrate on the correct word for one-half of the trials for each *S*.

the distribution, due in part to limiting each recognition series to 25 trials, a square-root transformation was performed. Analysis of these transformed data is shown in Table 1.

The analysis indicates that the groups differ significantly in the total number of trials required to emit the preselected response over all training frequencies ($F = 12.50$; $P < .01$). Except when a word appeared only once in pretraining, the ESP group consistently gave the appropriate response earliest and *S*s in the SLP group gave it latest in the series. The interaction between frequency of pretraining and groups is not significant. Both frequency of prior occurrence and instructions, therefore, have significant effects on the order of appearance of pretraining responses at each of the pretraining frequencies, and these effects are consistent for all the groups. Variation in the expected source of information, induced through instructions, had the predicted effects on "correct recognition."

Number of words recognized.—Comparison of number of words recognized represents another measure of the effects of experimental treatment. It has an advantage in that it is free of any biasing effect due to restricting each series to 25 trials. The mean number of words recognized by the respective experimental groups was: ESP, 5.71; LNG, 4.47; and SLP, 3.36. Figure 2 presents these data, converted to percentages, for the three groups at each of the training frequencies. The ESP group reaches 100% accuracy of "recognition" at the highest training frequency and also shows higher average percentage of words recognized at each of the frequencies, excepting one. Analysis of variance of the mean number of words recognized yields F ratios beyond the .01 level for both frequency ($F = 35.21$) and groups ($F = 10.47$). These results are consistent with those obtained in the analysis of trials to recognition.

TABLE 1

Analysis of Variance of Square Root Transformations of Trials to "Correct" Response for 10 Series of Preselected Words of Five Varying Frequencies

Source	*df*	*MS*	*F*
Between groups (G)	2	21.968	12.50*
Pooled *S*s in groups	39	1.757	
Within groups	378		
Training frequency (F)	4	82.227	72.57*
F × G	8	1.947	1.72
F × W *S*s pooled in groups	156	1.994	1.76*
Within-cells error	210	1.133	

* $P < .01$.

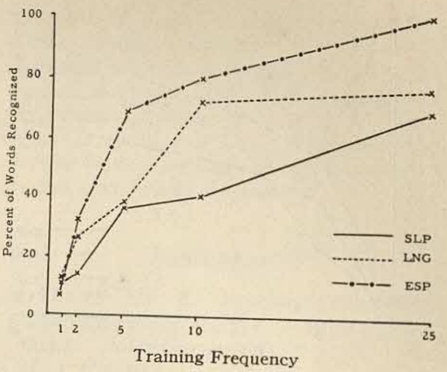


FIG. 2. Percentage of words recognized as a function of training frequency and instruction.

Number of words recalled.—If frustration and/or negative reinforcement associated with incorrect responses during the task series did not alter response biases, no differences among these groups would be expected in recall, since the "stimulus" conditions on this task were identical for all groups.[3] The average number of

---

[3] In contrast to the preceding task, all *S*s were instructed to "try to remember"; thus the recall condition was most similar to the LNG instructions.

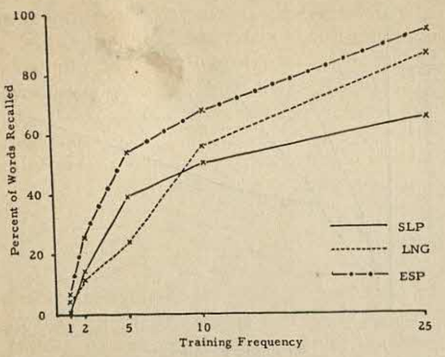


FIG. 3. Percentage of words correctly recalled as a function of training frequency and instructions.

correctly recalled words was 4.86 for the ESP group, 3.66 for the LNG group, and 3.37 for the SLP group. Analysis of variance of these data yielded an F of 4.57 ($P < .05$) for the main effect of groups.

Figure 3 presents the percentage of words correctly recalled as a function of training frequency and experimental treatment. The order of the groups on this measure is similar to that obtained on the other measures with the greatest effects of pretraining again reflected in the performance of the ESP group.

DISCUSSION

These results indicate that response biases in quasiperceptual situations may be affected not only by frequency of pretraining but also by response sets induced through instructions. More specifically, it was found that a perceptual set attenuates the dependency of "recognition" responses on frequency of prior experience even in the absence of a specified stimulus. These data lend support, therefore, to the hypothesized effects of variation in the expected source of information and consequences of response accuracy confirmation on response biases in perceptual task situations.

If behavior is conceptualized as a process in which perceptual responses mediate overt responses (Kanfer, 1956), then one would expect any increase in irrelevant external stimulation to evoke a greater number of perceptual identifying responses and, therefore, greater variability in overt responses. Further, such "attending responses," whether aroused by instructions or subsequent irrelevant stimulation, lead to informational input that may be incongruent with S's response biases. In the subliminal perception conditions, for example, random sensory input from the flash on the screen, etc. would, according to this assumption, evoke identifying responses which then modified and restricted the availability of pretrained responses. No such irrelevant events were systematically attended to in the other groups. However, instructions to the LNG group did emphasize that response accuracy was an important factor in the task. It was predicted that failure of confirmation of response accuracy would generate frustration and emotional concomitants of failure that would effectively change the biases established in the training session. That this factor alone is sufficient to modify response probability may be inferred from the performance of Group LNG. Finally, the performance of Group ESP obtained under conditions where neither the orientation toward external information nor response accuracy confirmation constituted aspects of the experimental treatments indicates that these instructional factors reduce the maximum effectiveness of prior training.

Omission of an E-defined stimulus is an insufficient basis for the inference that perceptual variables do not systematically influence indicator responses in perceptual tasks. The simple instruction to the human S that he attend to an external stimulus (a constant but not trivial aspect of perceptual research) produces altered states of the organism which must be taken into account whenever the effects of additional variables on perceptual processes are to be studied. If the definition of perception is thus broadened to include situational and organismic factors, such as those which influenced behavior in this experiment, then the quasiperceptual situation may serve as a method which will permit the isolation and exploration of some of the complex variables operating in perception tasks. The isolation of these component processes and their contribution to overt responses in various perceptual situations remains a primary goal of future research. For example, determination of motivational and other attitudinal factors operating to modify response biases may help to resolve further the contradictions arising from perceptual defense studies.

SUMMARY

The present study determined the effects of expected source of information on response biases in a quasiperceptual task. It was predicted that response biases, acquired through prior training, would contribute maximally to "recognition" responses under conditions where: (*a*) *S*s expected no additional external information; (*b*) accuracy of recognition was minimized. Each of 42 *S*s was pretrained by pronouncing nonsense words presented for differing number of trials. Following pretraining, *S*s were divided into three equal groups. Group ESP read interpolated material concerning extrasensory perception and was then instructed to respond to *E*'s "thought concentration" during the quasiperceptual task situation. Group LNG was given comparable reading material emphasizing subconscious factors in learning and memory and then was instructed to search his memory during the task situation. Group SLP read articles suggesting the effectiveness of subliminal perception and was instructed to respond to "subliminally" presented words during the task situation. No stimulus word was presented during the task situation to any of the groups. Accuracy of responses as measured by trials to recognition, number of words recognized, and recalled words, was significantly greater for Group ESP than either Groups SLP or LNG. The implications of these findings for methodology in perception studies were briefly discussed.

REFERENCES

BLACKWELL, H. R. Psychophysical thresholds: Empirical studies of their measurement. *Bull. Engng. Res. Inst. U. Mich.*, 1952, No. 36.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistiscopic stimuli convey some information? *Psychol. Rev.*, 1953, **63**, 181–188.

GARNER, W., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concepts of perception. *Psychol. Rev.*, 1956, **63**, 149–159.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between word frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

HOCHBERG, J. E. Perception: Toward the recovery of a definition. *Psychol. Rev.*, 1956, **63**, 400–405.

JAMES, W. H., & ROTTER, J. B. Partial and 100% reinforcement under chance and skill conditions. *J. exp. Psychol.*, 1958, **55**, 397–404.

KANFER, F. H. Perception: Identification and instrumental activity. *Psychol. Rev.*, 1956, **63**, 317–329.

NEISSER, U. An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, 1954, **47**, 399–402.

(Received June 10, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 164–170

REINFORCING PROPERTIES OF THE ONSET OF AUDITORY STIMULATION[1]

GERALD W. BARNES[2] AND GEORGE B. KISH

University of Maine

Recent reviews of the experimental literature (Glanzer, 1958; Waters, Rethlingshafer, & Caldwell, 1960) summarize a growing body of data which is congruent with the "sensory reinforcement" hypothesis, i.e., operant behavior probability may be increased and maintained by stimuli unrelated to the usual organic need states (Barnes, Kish, & Wood, 1959). The data reviewed appear to fall into four general groupings:

1. Studies reported by Harlow et al. (e.g., Harlow, 1950; Harlow, Harlow, & Meyer, 1950) report that monkeys will engage in extensive manipulatory and puzzle-solving behavior without the presentation of any of the ordinary primary or secondary reinforcers.
2. Butler reported excellent learning by monkeys if the "correct" response led to such visual stimulation as a 30-sec. exposure to the environment outside the test cage (Butler, 1953) or such auditory stimulation as the sounds arising from the monkey colony (Butler, 1957).
3. Montgomery has reported in a series of papers (e.g., Montgomery, 1954; Montgomery & Segall, 1955) that mazes may be learned by rats if the "correct" response leads to any one of a variety of complex stimuli (e.g., a Dashiell maze) which has never been experimentally associated with primary drive reduction.
4. The above groups of studies all involve stimulation of extreme complexity. A fourth group of studies deals with the effects of much less complex stimulation on relatively simple behavior. These studies uniformly report that the onset of dim illumination is intrinsically reinforcing for rats (e.g., Girdner, 1953; Marx, Henderson, & Roberts, 1955) and for mice (e.g., Barnes et al., 1959; Kish, 1955).

A basic question, relative to the sensory reinforcement hypothesis, raised by the fourth group of studies is that of the generality of the phenomenon. Will stimulation in modalities other than vision prove to be sensory reinforcers? The studies of Harlow, Butler, and Montgomery suggest that sensory reinforcement may operate in situations other than contacting or depressing a lever and that tactual, kinesthetic, and auditory stimuli may serve as sensory reinforcers. However, the extreme complexity of the stimuli used in these studies and the possibility of their previous association with primary reinforcement preclude unambiguous interpretation. For example in Harlow's studies of manipulation, visual, tactual, and auditory stimulation as well as kinesthetic stimulation may be operating concurrently. Also, Butler's (1957) finding that monkeys will learn a response which produces sounds from the monkey colony appears to be complicated by the secondary reinforcing properties assignable to such sounds through the past history of *S*s in the colony itself.

The objective of the present study, therefore, is to investigate the effects of various frequencies and intensities of auditory stimulation upon the lever-contacting behavior of mice in a situation in which response-contingent stimulation in other sensory modalities has been reduced to a minimum.

METHOD

Subjects.—A total of 1000 *S*s, 880 male and 120 female C57 Black Subline 6 mice, 50–70

---

[1] This research was supported by Research Grant NSF-G8609 from the National Science Foundation.

[2] Now at the University of Arkansas.

days of age were tested in this study. The *S*s were received in blocks of 120 and housed in a large communal living cage in which Purina lab chow and water were constantly available. Female *S*s were equally distributed among the 40 conditions of the second phase of the experiment.

Apparatus.—The apparatus has been extensively described elsewhere (Barnes et al., 1959) and comprised three identical sets of equipment, each consisting of a $5\frac{3}{4} \times 5\frac{3}{4} \times 5\frac{3}{4}$ in. ventilated Skinner-type box placed within a large, sound-insulated wooden box. Base noise level in the box was 30 db. From one of the flat black walls of the box an immovable U-shaped bar protruded $\frac{5}{16}$ in. Contacts by *S* with this bar activated a Hunter electronic relay which in turn operated cumulative recorders (Gerbrands Model C and/or ZDG-I 3/0 printing counters), Cramer .01-min. running time meters, and sound presentation equipment located in an adjoining room.

The top of each box was closed by a wooden framework containing a Quam speaker Type 732, covered by a wire screen, during the first phase of the experiment. During the second phase the wooden framework contained an Olson S-297 high-frequency tweeter. Sound stimulation was produced by an H. H. Scott Type 810-A white noise generator and Heathkit Audio Signal Generators, all amplified by Heathkit high fidelity amplifiers.

Control equipment provided for a completely automated experimental cycle 1 hr. in length, divided into three consecutive 20-min. experimental periods. Thus the three experimental periods were timed, data were recorded, and stimuli were presented automatically.[3]

Procedure.—The experiment utilized the $10 \times 5 \times 2$ factorial design defined by 10 frequencies, 5 intensities, and 2 contingencies, as follows: (*a*) Frequencies (in cps): white noise, 700, 1200, 2000, 4000, 6000, 9000, 12000, 14000, 16000. (*b*) Intensities (in db.): 45, 55, 65, 75, 85. (*c*) Contingencies: tone contingent on bar contact, tone not contingent on bar contact.

During the first 20-min. period (hereafter called "operant level") of the 1-hr. experimental session, the unconditioned rate of bar contacting was determined; no auditory stimulation was presented during this period. During the second 20-min. period (hereafter called "conditioning") the Contingent groups received pure tone or white noise of appropriate intensity and frequency as the consequence of each bar contact; duration of stimulation was equal to duration of bar contact. In the Noncontingent groups, sound stimulation of appropriate frequency and intensity was presented for 1 sec. every 18 sec., independently of the behavior of *S*. During the final 20-min. period (hereafter called "extinction") the operant level conditions, i.e., no sound stimulation, were reinstated.

Owing to technical problems related to sound production and sound intensity measurement, the experiment was run in two phases. Phase I included frequencies from white noise through 6000 cps and Phase II included frequencies from 9000 cps through 16000 cps. For measurements in Phase II, a calibrated microphone (Electrovoice Model 655C) was substituted for the original Rochelle salt microphone on the General Radio Sound Level Meter, thus requiring a slight change in the technique of sound measurement. All sound measurements referred to in the experiment was taken on the "A" weighting of the sound level meter.

During Phase I, *S*s were run in 10 squads of 60 mice each, randomly distributed among the 60 possible experimental conditions. Similarly, *S*s in Phase II were run in 10 squads of 40 mice each, randomly distributed among the 40 possible experimental conditions. Thus 100 independent groups of 10 *S*s each were tested.

Data consisted of (*a*) cumulative response records of the performance of each *S* during the entire experimental session and (*b*) total numbers and total durations of bar contacts during each 20-min. period of a given experimental session.

RESULTS

Cumulated response times in each experimental period were converted for each *S* into average time per response (hereafter called T), in seconds, for each period. The T measures and total numbers of responses emitted during the operant level, conditioning, and extinction periods were then converted by subtraction for each *S* into the difference scores (*a*) C-OL (Conditioning minus Operant Level) and (*b*) OL-E (Operant Level minus Extinction). This conversion was made because exten-

---

[3] Automation of the experiment was made possible by a grant-in-aid from the Wilson Coe Research Fund, University of Maine.

TABLE 1

F VALUES FOR ALL MEASURES

Source of Variation	*df*	Total Responses		Selected Responses[a]		T = Time/Responses	
		C-OL	OL-E	C-OL	E-C	C-OL	OL-E
Frequency (Fr)	9,900	4.14**	0.74	3.84**	1.02	0.94	1.34
Intensity (db.)	4,900	2.65*	1.52	2.18	1.29	5.23**	1.12
Contingency (Co)	1,900	0.52	1.32	0.15	2.03	1.59	0.04
Fr × Co	9,900	1.92*	1.14	0.62	2.48**	1.90*	1.28
db. × Co	4,900	1.49	0.69	0.27	0.37	3.14*	1.24
Fr × db.	36,900	1.30	0.93	1.20	0.84	1.44*	1.19
Fr × db. × Co	36,900	0.87	0.69	0.85	1.15	1.28	1.21

[a] Comparison is between responses in first 5 min. of one period (e.g., C) and last 5 min. of second period (e.g., OL).
* $P < .05$.
** $P < .01$.

sive experience with light-onset reinforcement has indicated that such difference measures are considerably more sensitive than are the raw data. The outcomes of analyses of variance of the difference scores are summarized in Table 1, and Fig. 1–4 display the data for those difference scores yielding at least one *F* value significant at or beyond the .05 level.[4]

The figures indicate that, in spite of the moderately large groups studied, considerable irregularity is the rule. Some general trends may, however, be noted. It may be seen in Fig. 1, for example, that only rarely did a group emit more responses during the conditioning period than during the operant level period; yet an overall downward trend is clearly apparent as frequency increases, thus accounting for the highly significant *F* value for this major variable. A

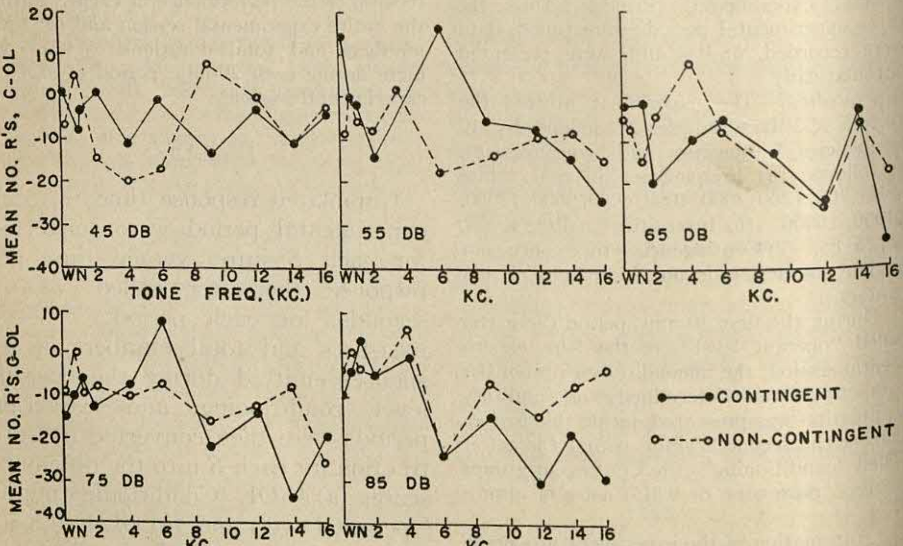


FIG. 1. Mean C-OL response measures for each group. (Each point represents the performance of an independent group of 10*S*s.)

---

[4] All computations were done at the MIT Computation Center, Cambridge, Massachusetts.

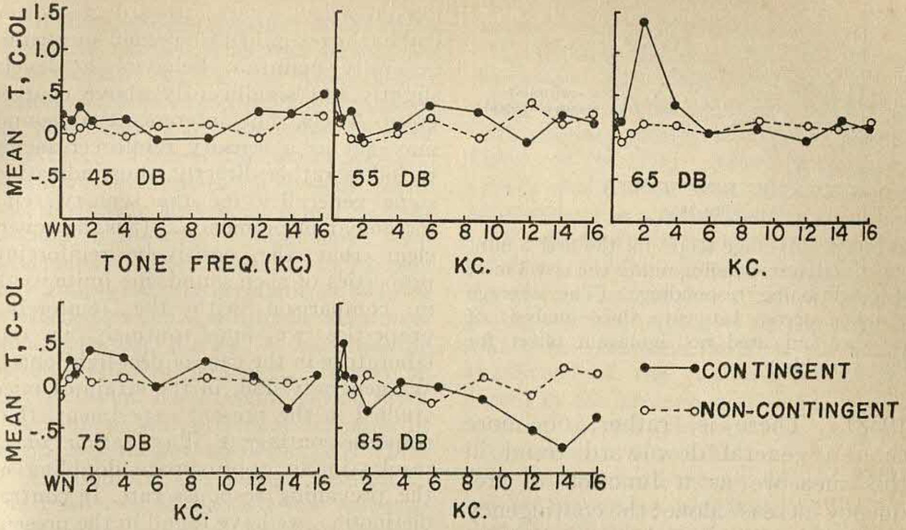


FIG. 2. Mean C-OL average time per response (T) for each group. (Each point represents the performance of an independent group of 10 *Ss*.)

significant inverse relationship between sound intensity and the C-OL measure for responses is also present, although this function is probably not a maximum at the lowest intensity of 45 db. The significant Frequency × Contingency interaction reflects a trend for the contingent groups to be above their noncontingent controls at frequencies below 6000 cps but below the noncontingent groups at higher frequencies.

Figure 2 indicates that most groups, both contingent and noncontingent, yielded higher average response times during the C period than during the OL period. The *F* values reported in Table 1 indicate, however, that the conclusions reached in the preceding paragraph concerning response frequencies also follow, perhaps even more clearly, from the T data. The significant Intensity × Contingency interaction for the C-OL difference for T is apparently due to a general trend for the contingent groups to yield larger T values than their controls at intensities of 75 db. and below, but smaller T values above 75 db.

Difference scores obtained by subtracting numbers of responses made during the last 5 min. of OL from those made during the first 5 min. of C are displayed in Fig. 3. This measure does not show the abrupt acceleration in response rate characteristic of contingent groups in light-onset experiments when the conditioning period begins (Barnes & Kish,

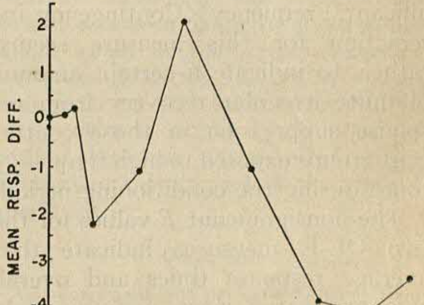


FIG. 3. Average scores for the first 5 min. of conditioning responding minus the last 5 min. of operant level responding. (The average is taken across Intensity and Contingency since analysis of variance indicated no significant effects for those two variables.)

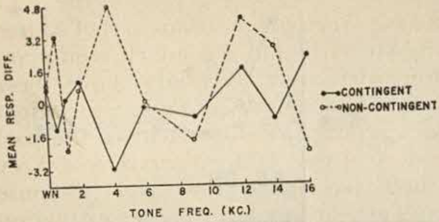


FIG. 4. Average scores for the first 5 min. of extinction responding minus the last 5 min. of conditioning responding. (The average is taken across Intensity since analysis of variance indicated no significant effect for that variable.)

1958). There is, rather, no more than a general downward trend in this measure as a function of frequency increase alone; the contingency variable appears to play no role here.

An acceleration is sometimes found at the beginning of extinction following a conditioning period with the usual organically-based reinforcers (e.g., Skinner, 1938). However, scores obtained by subtracting numbers of responses made during the last 5 min. of conditioning from those made during the first 5 min. of extinction (see Fig. 4) do not show that a general acceleration in response rate occurred when the extinction period began in the present experiment. The significant Frequency × Contingency interaction for this measure seems, rather, to indicate a certain amount of quite irregular recovery from response suppression in those contingent groups exposed to high frequency tones during the conditioning period.

The nonsignificant *F* values for the two OL-E measures indicate that average response times and overall response rates were approximately equal for all groups during the operant level and extinction periods.

DISCUSSION

The time and frequency data apparently indicate that response-contingent low-intensity sounds, toward the lower end of the examined frequency spectrum, generally maintain behavior at levels slightly but significantly above control group levels. The inference that sound may act as a sensory reinforcer seems to follow rather directly, thus indicating some generality for the sensory reinforcement phenomenon. It is, however, clear that the positively reinforcing properties of such sounds are quite weak in comparison with the reinforcing properties we find routinely in our laboratory in the case of dim-light onset. We generally find, in the strain of mice studied in the present experiment, that response-contingent illumination onset results in an approximate doubling of the prevailing response rate. In contradistinction, we have found in the present study that response-contingent tones do not in general increase the prevailing rate at all. Rather, response-contingent tones, in certain ranges, simply maintain the rate at an overall level greater than would be expected if such tones were not response-contingent. In the low frequency and low intensity ranges, then, tone onset seems to operate as a weak positive reinforcer, with reinforcing strength especially indicated by the occasional marked peaks in the C-OL measures for contingent groups. The existence of peaks of such magnitudes at specific frequency and intensity values would seem, however, to require independent replication before definite conclusions concerning them can be reached.

At frequencies above 9,000 cps the weak positive reinforcing effects of tone onset were apparently replaced by an aversive effect rather clearly exhibited by the 75-db. and 85-db. contingent groups. It is interesting to note that 85-db. tones contingent upon bar contact did not produce a response suppression over the total frequency range here investigated; such suppression was not evident until frequencies of 12,000 cps and above were involved. To what extent this suppression-failure depends upon the auditory capabilities of the mouse is a matter for speculation, for these capabilities do not seem to be known.

Although the positively reinforcing effects of the experimental treatments were evident during the conditioning period, such effects were not sufficient to yield significant response or time differences between the extinction and operant level periods. The same result has, however, also been found by the authors in the case of light-onset reinforcement (Barnes & Kish, 1958), and may simply be characteristic of sensory reinforcers when continuous reinforcement prevails during a conditioning period of 20 min. or longer. It may be that a 20-min. conditioning period, with a continuous reinforcement schedule in effect, is sufficient time for production of satiation for the sensory reinforcer. The result would be that low response rates appear during an immediately subsequent extinction session, and extinction responding would then not differ significantly from operant level response measures. More data on this point are clearly needed, for massive carry-over into extinction of the effects of a reinforcement period are easily demonstrated for such reinforcers as food and water.

One hypothesis which might be suggested immediately to account for the weakly positive effects of tones at low frequencies and intensities but negative effects at higher ones would be that, in the case of tones at least, positively reinforcing properties are exhibited at frequency values toward the limits of the hearing range of the mouse. The data presented herein would then be accounted for as being toward the bottom of the hearing range of the mouse, and one would then wish to examine frequencies toward the upper limit of the range. Verification of such an interpretation must await the collection of absolute threshold data, as a function of frequency, for the mouse. Data of this nature are not available, to the knowledge of the authors.

Another hypothesis arises from the everyday observation that moderate response rates are generated in human children by toys and other environmental objects which can be made to produce noises. Such observations are, of course, unsystematic and secondary reinforcing properties may certainly have been assigned to sounds by their associations with primary reinforcement in the past history of the child. Nevertheless, Frey (1960) has found that high response rates appear in a lever-pressing situation with preschool children when bar pressing produces a meaningless sound. It is, therefore, conceivable that a phylogenetic variable operates in the case of sounds-as-reinforcers, so that the small magnitude of the positive effects reported in the present experiment is due in part to the phylogenetic position of the mouse. Mammals better adapted than the mouse to the use of sounds as cues for movement might display a more marked sensory reinforcement effect than we were able to demonstrate with mice.

A recent experiment from our laboratory (Barnes & Baron, 1961) suggests a third hypothesis. This experiment involved substituting for the light reinforcer which we have always used with mice (i.e., a diffuse overall illumination of the entire experimental compartment) another visual stimulus, viz., a small pattern displayed on a Lucite screen just above the response lever. It was found that different patterns yielded response rates which differed at unusually great levels of statistical significance. An obvious extension of the result to sound stimuli would be that response-contingent sound patterns, perhaps both temporal patterns and tones presented in chord patterns, might be expected to produce response rates substantially higher than those found in the present experiment.

SUMMARY

An experiment was performed, using 1000 mice as *S*s, to test for possible reinforcing effects of the onset of various frequencies and intensities of auditory stimulation. The experiment utilized a factorial design defined by 10 frequencies, 5 intensities, and 2 contingencies (sound contingent on bar contact and sound not contingent on bar contact). Each *S* was tested in a single 1-hr. Experimental session consisting of a 20-min. Operant

Level period, followed by a 20-min. Conditioning period, followed by a 20-min. Extinction period. Both total number and total duration of bar contacts were recorded during each of the 20-min. periods.

The results indicated that auditory stimulation, in certain frequency and intensity ranges, showed positive reinforcing properties primarily by maintaining response rates in the contingent groups above those of the noncontingent groups during the conditioning period. This effect varied as a roughly inverse function of both frequency and intensity. The data were interpreted as showing that onset of auditory stimulation can function as a sensory reinforcer, though a considerably weaker one than onset of dim illumination.

REFERENCES

Barnes, G. W., & Baron, A. Stimulus complexity and sensory reinforcement. *J. comp. physiol. Psychol.*, 1961, **54**, 466–469.

Barnes, G. W., & Kish, G. B. On some properties of visual reinforcement. *Amer. Psychologist*, 1958, **13**, 417. (Abstract)

Barnes, G. W., Kish, G. B., & Wood, W. O. The effect of light intensity when onset or termination of illumination is used as reinforcing stimulus. *Psychol. Rec.*, 1959, **9**, 53–60.

Butler, R. A. Discrimination learning by rhesus monkeys to visual exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46**, 95–98.

Butler, R. A. Discrimination learning by rhesus monkeys to auditory incentives. *J. comp. physiol. Psychol.*, 1957, **50**, 239–241.

Frey, R. B. The effects of verbal reinforcers on group operant behavior. Unpublished master's thesis, University of Maine, 1960.

Girdner, J. B. An experimental analysis of the behavioral effects of a perceptual consequence unrelated to organic drive states. Unpublished doctoral dissertation, Duke University, 1953.

Glanzer, M. Curiosity, exploratory drive, and stimulus satiation. *Psychol. Bull.*, 1958, **55**, 302–315.

Harlow, H. F. Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *J. comp. physiol. Psychol.*, 1950, **43**, 289–294.

Harlow, H. F., Harlow, M. K., & Meyer, D. R. Learning motivated by a manipulative drive. *J. exp. Psychol.*, 1950, **40**, 228–234.

Kish, G. B. Learning when the onset of illumination is used as reinforcing stimulus. *J comp. physiol. Psychol.*, 1955, **48**, 261–264.

Marx, M. H., Henderson, R. L., & Roberts, C. L. Positive reinforcement of the bar-pressing response by a light stimulus following dark operant pretests with no aftereffect. *J. comp. physiol. Psychol.*, 1955, **48**, 73–75.

Montgomery, K. C. The role of exploratory drive in learning. *J. comp. physiol. Psychol.*, 1954, **47**, 60–69.

Montgomery, K. C., & Segall, M. Discrimination based on the exploratory drive. *J. comp. physiol. Psychol.*, 1955, **48**, 225–228.

Skinner, B. F. *Behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

Waters, R. H., Rethlingshafer, D. A., & Caldwell, W. E. *Principles of comparative psychology.* New York: McGraw-Hill, 1960.

(Received June 11, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 171–178

DIFFERENTIAL EFFECTS OF SHOCK IN HUMAN MAZE LEARNING[1]

SAMUEL M. FELDMAN[2]

Allan Memorial Institute, McGill University

The early observation of Tolman, Hall, and Bretnall (1932) that electric shock facilitates learning, even when the shocks follow correct responses, has been confirmed by an impressive number of subsequent investigations (e.g., Feldman, 1954; Freeburne & Schneider, 1955; Freeburne & Taylor, 1952; Muenzinger, 1934; Muenzinger, Bernstone, & Richards, 1938; Muenzinger, Brown, Crow, & Powloski, 1952; Muenzinger & Newcomb, 1935; Muenzinger & Powloski, 1951; Prince, 1956). Muenzinger and Baxter (1957), on the basis of an investigation of the perceptual properties of shocks following correct (shock-right) and incorrect (shock-wrong) responses, have suggested that two factors operate when *S*s are shocked. First, shock operates as a general facilitating mechanism, regardless of whether administered for correct or for incorrect responses. This is opposed to the common assumption that the only effect of shock is to produce avoidance. In addition to this, according to Muenzinger and Baxter, the shock takes on cue properties, leading to approach behavior in shock-right groups and to aversive behavior in shock-wrong groups. The shock-right groups, however, must overcome an initial aversion to shock. It is this initial aversion that accounts for the ultimate superiority of shock-wrong over shock-right in Muenzinger's data.

In this explanation, Muenzinger and Baxter make two assumptions. The first is that shock-wrong groups always learn faster than do shock-right groups, an assumption which is not borne out by the available evidence. While this is usually the case, several experiments have resulted in shock-right being the superior condition for learning (Feldman, 1954; Tolman et al., 1932), and in still others, no difference has been observed between shock-right and shock-wrong groups (e.g., Freeburne & Schneider, 1955). A review of earlier studies suggests that intensity of the noxious stimulus employed may be the basis of these conflicting findings (see e.g., Muenzinger & Newcomb, 1935). It seems that when shock has been held to a relatively mild intensity (e.g., by allowing human *S*s to adjust the intensity until they can "take no more"), shock-right has been the superior condition; but when the intensity of the shock has been relatively great (as in animal studies in which, e.g., *E* adjusts the current until *S* "jumps"), shock-wrong groups tended to perform better.

---

[1] This paper is based on a thesis submitted to the Faculty of Graduate Studies and Research, McGill University, in partial fulfilment of the requirements for the PhD degree. Support for the investigation has come from: Medical Research and Development Division, Office of The Surgeon General, Department of the Army, Contract DA-49-007-MD-626; National Institute of Mental Health, United States Public Health Service, Grant M-1475; Department of National Health and Welfare (Canada), Grant 604-5-69; Defence Research Board of Canada, Grant 9425-04; and National Research Council of Canada, Grant A.P. 29. The author is indebted to R. B. Malmo for his supervision, and for his generous provision of laboratory facilities. He also wishes to thank W. R. D. Ross for his technical assistance, and M. M. Schnore for his many constructive comments and criticism.

[2] Now at the Department of Physiology, Albert Einstein College of Medicine, New York City.

The second assumption of Muenzinger and Baxter is that cue properties alone differentiate the two shock conditions (i.e., shock-right vs. shock-wrong). If intensity is a critical factor in determining the relative effects of the two shock conditions, then the arousing properties of the shock stimulus bear investigation as well as the cue properties.

The present study was designed to obtain data bearing on questions concerning the interaction between shock intensity and shock condition in learning. To this end, an objective criterion of the arousing properties of the shock stimulus was required. Several authors (e.g., Duffy, 1957; Malmo, 1958; Schlosberg, 1954) have recently suggested that the measurement of the level of activity in certain physiological systems might provide meaningful indicants of arousal. Experiments carried out on human and animal *S*s at McGill University and at the University of Montreal have lent concrete support to this suggestion (see Malmo, 1959). The measurement of respiration rate, palmar conductance, and heart rate was therefore undertaken in the present investigation in order to study the arousing properties of shock intensity and shock condition in relation to the cue properties of these variables in a learning situation.

METHOD

Design.—The *S*s were randomly assigned to one of four groups, 15 to a group, in a 2×2 factorial design. All *S*s received 15 trials on a 25-choice point "nailhead" finger maze. Half of the *S*s received shocks for incorrect responses while the other half were shocked for correct responses. Each of these two shock conditions were divided into a "high" intensity and a "low" intensity group. The four groups were thus high shock-right (HSR), low shock-right (LSR), high shock-wrong (HSW) and low shock-wrong (LSW).

Subjects.—The *S*s were 60 male undergraduates at McGill University. They were paid for their services at the rate of $1.00 per hour.

Apparatus.—A 25-choice point nailhead maze was constructed by imbedding 50 copper nails in a plywood board. They formed two adjacent columns of 25 nails each, which appeared to *S* as a mounted board containing a single vertical column of 25 horizontal pairs of nailheads. The correct and incorrect sequences of nailheads were wired in parallel, independently of each other. This made possible shock administration and the recording of all responses. Each adjacent pair of nailheads thus constituted a single choice point. The correct sequence, for all groups, was L L R L R L L R R R L R R L L L R R L R R L L R L.

When *S* touched a nailhead which was in the shocking circuit a condenser was discharged to his body through a large series resistance, which served to minimize individual differences in resistance. A shielded cable extended the shocking circuit from a control chamber to *S*'s room, where one side was connected to the maze (either to the correct or to the incorrect sequence) and the other side was connected to a strip of brass, which was wrapped around the proximal phalanx of *S*'s right index finger. A metal thimble, filled with electrode jelly, was placed on this finger. When the thimble was brought into contact with the shock-wired nailheads, the circuit was completed through *S*'s finger. Current flow, at the moment of discharge of the condenser was 3 ma. for the low-shock groups and 9 ma. for the high-shock groups.

Testing procedure.—After a brief interview in *E*'s office, *S* was taken into the testing chamber and seated in a chair with adjustable arm rests. All electrodes and attachments were applied, except for the shocking electrodes. The testing room was a shielded chamber which contained a table, adjustable chair, and electrodes, lead wires, and chemicals necessary for electrode attachment. An adjacent room, with a window looking into the testing chamber, contained all of the recording equipment. During the experiment, *S* was observed by *E* through the window, and *S* could be heard through an intercommunication system.

In order to get an estimate of *S*'s level of physiological activity in a relatively relaxed condition, *S* was initially told that the study was a comparison of physiological activity during reading and during relaxation. After a 25-min. period of magazine reading, with recording apparatus attached, and equipment turned on, *S* was asked to relax. A 5-min. record of palmar resistance, respiration rate, and heart rate recorded during this session was used as a personal baseline for *S*. It was after this record had been taken that the learning task was introduced. Instructions for learning were given, the maze was placed on the table before *S*, and shocking leads were applied. Each *S* had 15 trials, with a 1-min. interval between trials. A trial consisted of a single traversal of the complete maze, choosing one nailhead in each of the 25 pairs. No practice trials were given.

Recording technique.[3]—Heart rate was recorded by means of a continuous electrocardiographic tracing taken on an Offner Type D electroencephalograph. A chest placement was used for the electrodes, which were made of cellulose sponge, saturated with normal saline and electrode jelly. This Offner output signal was recorded on a chart drive which contained tracings of this and other systems which follow, as well as the responses made in learning.

The period selected for sampling physiological activity was the pretrial period. The experiment required an estimate of *S*'s level of arousal produced by the total situation. Since a specific reaction-to-shock measure was not desired, the posttrial period was not selected for study. The pretrial measurements were samples of activity occurring during the period roughly 10 sec. prior to each trial. Since the four groups were independent, absolute level of activity could not be used. An estimate of *S*'s resting level of activity had to be taken in order that *S*'s pretrial activity could be expressed in terms of a personal baseline. The resting level was estimated on the basis of pre-experimental samples, taken during the 5-min. resting period prior to the introduction of the learning task.

In the case of heart rate, the pre-experimental measure was an average of five 10-cycle (i.e., 11 beats) samples taken at the end of each of the five minutes during the pre-experimental "rest" period. The pretrial measure was an average of the 10 cycles preceding the beginning of each trial. These averages were based on conversions to beats per minute (b/m).

Palmar conductance was calculated from a continuous recording of palmar resistance. The circuit used was a standard comparator bridge, with a current varying no more than 2% when *S*'s resistance was in the range of 0 to 80,000 ohms. The monopolar technique employed for recording resistance has been reported in detail by Malmo and Davis (1960).

Palmar resistance was measured in ohms and then converted to conductance in micromhos. The measurements were made at the beginning and end of each of the 10-cycle heart rate samples. There were thus 10 measurements made during the pre-experimental period and two prior to each trial.

Respiration rate was determined from a continuous recording taken with a tambour and pneumograph. An abdominal placement was used for the accordion pneumograph. The method used for sampling respiration was the same as that used for sampling heart rate, except that three complete inspiration-expiration samples were measured rather than the 10-cycle sample used in the case of heart rate. These measurements were converted to respirations per minute (r/m).

In order to check on the reliability of these measurements, odd-even rank order correlations were calculated for each physiological measure, in each experimental condition. These were based on an average of each *S*'s activity on odd trials with activity on even trials. In all cases except one, the resultant rho was greater than +.95. The one exception was respiration rate in the HSR group, in which case rho was +.88.

Treatment of data.—Since *S*s received no practice trials, the number of errors made on Trial 1 was determined by chance alone. The error score used, therefore, was total errors/errors on Trial 1. In this way, each *S*'s score was a function of an individual baseline. The same technique was employed in calculating time scores, i.e., total time/time on Trial 1.

Each measure of physiological activity for every *S* was expressed as mean pretrial activity divided by mean pre-experimental activity. Absolute level scores were not appropriate since the groups were independent. The mean pre-experimental scores were based on the mean of five samples taken during the "resting" period.

It was felt that a measure of arousal should be able to differentiate between the high-shock *S*s and low-shock *S*s. Since this

---

[3] In addition to the systems reported here, electromyograms were recorded from the superficial extensor muscles of the left (inactive) forearm. These measurements were taken for purposes other than those of the present study, and are not discussed in this paper.

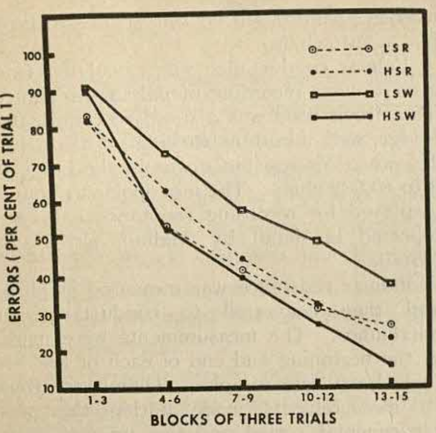


FIG. 1. Errors (percentage of Trial 1) as a function of practice.

has not been demonstrated for the measures employed in the present study, the criterion for the adoption of measures of arousal was their ability to provide this differentiation.

RESULTS

Learning.—As can be seen in Fig. 1, Group HSW made the least errors, Group LSW the most, while the two shock-right groups fell in between. This was the case for the time data as well, as is illustrated in Fig. 2. Table 1 presents a 2 × 2 chi square test of analysis of variance hypotheses (Wilson, 1956) for these data. In the case of both time and errors, neither shock intensity nor shock condition had a statistically significant effect; only the interaction term was significant. Mann-Whitney *U* tests (Table 2) showed that the difference between the LSR and LSW groups was significant for time and errors ($P < .05$), LSR being the condition associated with superior performance. Differences between Groups HSR and HSW were not statistically significant. When the mean time and error scores on Trials 11 to 15 were analyzed (i.e., Trials 11–15/Trial 1), the time data showed a clear reversal as a function of shock intensity, Groups LSR and HSW being significantly faster than Groups LSW and HSR, respectively. Differences in this analysis for errors were in the same direction, but failed to meet statistical significance.

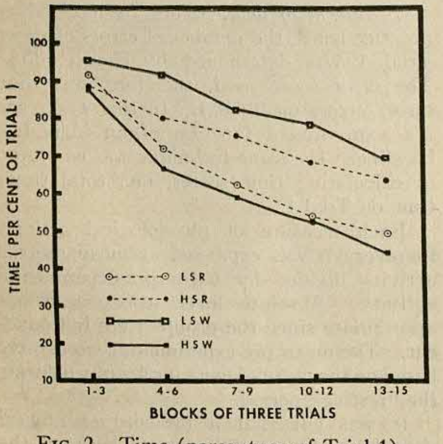


FIG. 2. Time (percentage of Trial 1) as a function of practice.

TABLE 1

2 × 2 CHI SQUARE ANALYSIS OF THE EFFECTS OF SHOCK INTENSITY AND SHOCK CONDITION ON LEARNING (TRIALS 1–15/TRIAL 1)

Source of Chi Square	Chi Square Errors	 Time
Total	11.7***	8.8**
Shock condition	2.4	1.1
Shock intensity	0.0	1.1
Interaction	9.3**	6.7**

Note.—$N = 15$ *Ss* per group.
** $P < .01$.
*** $P < .001$.

TABLE 2

U'S BASED ON DIFFERENCES BETWEEN GROUPS IN LEARNING

Groups Compared	Trials	*U* Time	 Errors
LSR-LSW	1–15	64.0*	63.5*
HSR-HSW	1–15	70.0	89.0
LSR-LSW	11–15	46.0**	74.0
HSR-HSW	11–15	62.0*	74.0

Note.—$N = 15$ *Ss* per group.
* $P < .05$.
** $P < .01$.

TABLE 3

MEAN LEVELS OF PHYSIOLOGICAL ACTIVITY

Group	Meas.	Respiration Rate (Resp/Min)			Palmar Conductance (Micromhos)			Heart Rate (Beats/Min)		
		(A) Mean Pre-Exp.	(B) Mean Pre-Trial	(C) B/A × 100	(A) Mean Pre-Exp.	(B) Mean Pre-Trial	(C) B/A × 100	(A) Mean Pre-Exp.	(B) Mean Pre-Trial	(C) B/A × 100
LSR	Mean	17.1	17.7	103.4	145.6	146.5	104.6	80.9	82.4	102.0
	SD	2.2	2.2	7.6	47.4	38.3	22.4	10.7	11.4	5.6
LSW	Mean	16.3	18.2	112.7	149.7	154.3	103.4	76.6	79.3	103.7
	SD	2.9	2.9	13.0	35.4	38.5	12.9	9.8	10.2	7.7
HSR	Mean	16.7	18.7	112.8	144.6	163.2	118.3	78.1	81.7	104.7
	SD	3.0	2.8	10.8	49.2	33.7	26.9	12.5	14.2	7.8
HSW	Mean	16.7	18.9	116.6	126.8	158.3	130.2	76.6	77.0	100.9
	SD	3.8	3.1	21.1	34.1	27.1	27.5	9.6	9.6	7.9

Note.—*N* = 15 *S*s per group.

In looking at the effect of increasing shock intensity, no significant differences were observed for time or for errors between the two shock-right groups. Group LSW was, however, consistently poorer in learning than Group HSW; all differences, for time and for errors, were statistically significant.

Physiological activity.—Table 3 contains the mean levels of physiological activity associated with the experimental conditions. Each entry in Column (C) represents a mean for 15 *S*s of mean pretrial/pre-experimental samples. It was pointed out earlier that any measure of physiological activity which is to distinguish between levels of arousal was required to differentiate the high-shock *S*s from the low-shock *S*s. As can be seen in Table 4, two of the three measures employed met this criterion; in the case of heart rate, there was no difference between the high-shock and low-shock groups.[4] In the case of respiration rate the difference between the high-shock and low-shock groups was significant beyond the .01 level, high shock being associated with faster respiration than low shock. It can be seen in Table 4 that the interaction term for respiration rate was also statistically significant; this is

TABLE 4

2 × 2 CHI SQUARE ANALYSIS OF EFFECTS OF SHOCK INTENSITY AND SHOCK CONDITION ON CHANGES IN LEVEL OF PHYSIOLOGICAL ACTIVITY

Source of Chi Square	Respiration Rate	Palmar Conductance	Heart Rate
Total	12.00***	5.60*	1.33
Shock condition	1.07	1.07	0.27
Shock intensity	6.67**	4.27*	0.00
Interaction	4.26*	0.99	1.06

Note.—*N* = 15 *S*s per group.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

---

[4] Malmo and Surwillo (1960), in a recent study of sleep deprivation, found that heart rate increased significantly over a 3-day vigil in two *S*s tested, but that heart rate decreased significantly in a third *S*. Activity in other physiological measures suggested that all three *S*s increased in overall level of arousal as a result of the vigil. Heart rate, thus, may be an indicant of arousal through an increase in some *S*s, and through a decrease in others. Further study, both of *S*s and experimental conditions in relation to heart rate, is needed.

due to the fact that the low-high difference was much greater in the case of shock-right than in the case of shock-wrong (Table 3).

Palmar conductance also met the above cited criterion. High shock was associated with greater conductance than was low shock. Table 4 shows this difference to be significant beyond the .05 level. Only heart rate failed to meet this criterion, and was therefore rejected as a measure of level of arousal in studying the shock conditions.

Wilson's 2 × 2 chi square test of analysis of variance hypotheses showed that the differences between shock conditions failed to meet statistical significance. The direction of the differences showed shock-wrong to be associated with higher levels of palmar conductance and respiration rate than shock-right. But these differences were not significant.

DISCUSSION

The many experiments in which humans and rats have been shocked for correct responses in learning have demonstrated that shock can act as a general facilitating mechanism. This is consistent with recent statements on the role of arousal in behavior (see Hebb, 1955; Malmo, 1959; Schlosberg, 1954). Muenzinger and Baxter (1957), in their investigation of the perceptual properties of shock, argued that shock-right learning could not be accounted for simply in terms of anxiety reduction (Mowrer, 1950; Prince, 1956). The present experiment represents an investigation of the arousing properties of the shock stimulus, in relation to the work of Muenzinger and his colleagues.

In an early paper, Muenzinger, Bernstone, and Richards (1938) spoke of the operation of both general and specific mechanisms in the use of shock. They assumed that more than one mechanism was operating, and that the specific mechanism (which Muenzinger and Baxter later concluded was primarily perceptual) did not contribute to the general one in an additive fashion. The data of the present experiment are consistent with this view, in that shock intensity was seen to be accompanied by differences in level of physiological activity, while differences in shock condition were not.[5]

In their report, however, Muenzinger and Baxter assumed that shock-wrong was consistently a better condition for learning than was shock-right. The data of our experiment show, on the contrary, that shock-right was the superior condition for learning when the shock intensity was relatively low. When shock intensity was relatively high, this was not the case; some measures of learning showed no significant difference, and others showed shock-wrong to be the superior condition. In all cases, the direction of difference favored shock-right at low intensity and shock-wrong at high intensity.

Muenzinger and Baxter further assumed that once the initial aversion to shock was overcome by shock-right *Ss*, they would then regard the shock, and all cues associated with it, as cues for approach behavior. In addition to this, they felt that this would serve to equalize the two conditions, and no differences would then be observed between shock-right and shock-wrong groups. Our data suggest that once the initial aversion to shock is overcome, shock-right, having the advantage of the shock and cues associated with it as signals for approach behavior, becomes the better condition for learning. This happened early in the case of Group LSR, but late in the case of Group HSR, resulting in the interaction between shock condition and shock intensity in learning. (Actually this failure to overcome the initial aversion to shock was quite varied in Group HSR, as was evidenced by the fact that the variance in this group, both for time and for errors, was significantly greater than for any other group.)

---

[5] A note of caution must be introduced with respect to the lack of significance of difference in respiration rate and palmar conductance between the shock-right and shock-wrong groups. All groups were independent, which is not an optimal condition for the study of differences in physiological activity due to experimental conditions, since the variance from *S* to *S* is considerable in these measures. The large variances in the present study could, conceivably, have led to a Type II error.

The data suggest, in summary, that the increased facilitating effect of increased shock, evident in the shock-wrong groups, was not evident in the shock-right groups due to the fact that the initial aversion to shock was overcome sooner in Group LSR than in Group HSR.

The present study thus supports the suggestions of Muenzinger and Baxter on the dual function of the shock stimulus. The interaction in the learning data, however, leads to a further conclusion: namely, that shock-right is the superior condition for learning after the initial avoidance tendencies are overcome, and it is perceived as a signal for approach behavior. This occurs sooner in the case of the low-shock *Ss* than in the case of the high-shock *Ss*.

SUMMARY

The main purpose of the present study was to investigate the role of shock in learning with reference to the problem of shock-right vs. shock-wrong. Four independent groups of 15 *Ss* each learned a "nailhead" finger maze. Shock intensity ("low" and "high") and shock condition ("right" and "wrong") were varied in a 2 × 2 factorial design. The results were consistent with previous findings in showing that shock-right was the superior condition when the intensity was low, but that shock-wrong tended to be the superior condition when the intensity of the shock was relatively high.

Continuous recordings of respiration rate and palmar conductance served to gauge the level of arousal produced by the experimental conditions. There was a significant difference in level of physiological activity accompanying the different intensities of shock in the present study, high shock being associated with higher levels of activity than low shock. No significant differences were found between physiological activity associated with shock-right (shock intensity constant) and that associated with shock-wrong. It appears, therefore, that the differential effects of shock on these two conditions cannot be accounted for in terms of an inherent difference in arousal level between the two conditions. Rather, it is suggested that the differential effects be interpreted in terms of Muenzinger and Baxter's suggestion that shock-right and shock-wrong are accompanied by differences in the perceptual role of shock.

REFERENCES

DUFFY, E. The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, 1957, **64,** 265-275.

FELDMAN, S. M. Escape from punishment and motivation in maze learning. Unpublished honor's thesis, University of Pennsylvania, 1954.

FREEBURNE, C. M., & SCHNEIDER, M. Shock for right and wrong responses during learning and extinction in human subjects. *J. exp. Psychol.*, 1955, **49,** 181-186.

FREEBURNE, C. M., & TAYLOR, J. E. Discrimination learning with shock for right and wrong responses in the same subjects. *J. comp. physiol. Psychol.*, 1952, **45,** 264-268.

HEBB, D. O. Drives and the C. N. S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62,** 243-254.

MALMO, R. B. Measurement of drive: An unsolved problem in psychology. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1958. Pp. 229-265.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66,** 367-386.

MALMO, R. B., & DAVIS, J. F. A monopolar method for recording palmar conductance. Montreal: Allan Memorial Institute of Psychiatry, 1960. (Mimeo)

MALMO, R. B., & SURWILLO, W. W. Sleep deprivation: Changes in performance and physiological indicants of activation. *Psychol. Monogr.*, 1960, **74**(15, Whole No. 502).

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

MUENZINGER, K. F. Motivation in learning: I. Electric shock for correct responses in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17,** 267-277.

MUENZINGER, K. F., & BAXTER, L. F. The effect of training to approach vs. training to escape from electric shock upon subsequent discrimination learning. *J. comp. physiol. Psychol.*, 1957, **50,** 252-257.

MUENZINGER, K. F., BERNSTONE, A. H., & RICHARDS, L. Motivation in learning: VIII. Equivalent amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.

MUENZINGER, K. F., BROWN, W. C., CROW, W. J., & POWLOSKI, R. F. Motivation in learning: XI. An analysis of electric shock for correct responses into its avoidance and accelerating components. *J. exp. Psychol.*, 1952, **43**, 115–119.

MUENZINGER, K. F., & NEWCOMB, H. Motivation in learning: III. A bell signal compared with electric shock for right and wrong responses in the visual discrimination habit. *J. comp. Psychol.*, 1935, **20**, 85–93.

MUENZINGER, K. F., & POWLOSKI, R. F. Motivation in learning: X. Comparison of electric shock for correct turns in a correction and non-correction situation. *J. exp. Psychol.*, 1951, **42**, 118–124.

PRINCE, A. I., JR. Effect of punishment on visual discrimination learning. *J. exp. Psychol.*, 1956, **52**, 381–385.

SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

TOLMAN, E. C., HALL, C. S., & BRETNALL, E. P. A disproof of the law of effect and a substitution of the laws of emphasis, motivation, and disruption. *J. exp. Psychol.*, 1932, **15**, 601–614.

WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

(Received June 13, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 179–186

RESPONSE LATENCY AS A FUNCTION OF SIZE OF GAP IN THE ELEVATED RUNWAY

DAVID BIRCH, L. THOMAS CLIFFORD,[1] AND JULIE BUTTERFIELD[2]

University of Michigan

Motivated rats required to locomote from an approach platform to a goal platform to obtain a goal object may be expected to vary in their response latency as a function of the extent of the gap separating the two platforms. The present study includes two experiments designed to yield information relevant to the theoretical interpretation of the effect of the size of gap on response latency. In Exp. I, three experimental manipulations, time of water deprivation (T_D), size of gap (G), and amount of training (N) are imposed and the pattern of interactions arising from the effects of these three manipulations, $E(T_D)$, $E(G)$, and $E(N)$, is under investigation. In Exp. II, an implication of the hypothesis that the gap elicits responses incompatible with the approach to the goal platform is tested.

Experiment I

Each *S*, under either 20 or 4 hr. water deprivation, crossed both a large and small gap to water an equal number of times each day, and this combination was repeated on each of 18 successive days with response latency (L) as the measure. This procedure results in a three dimensional factorial design of Lindquist's Type VI (Lindquist, 1953).

Method

Subjects.—The *S*s were 30 female pigmented rats from the University of Michigan Psychology Department colony, approximately 260 days old at the beginning of the experiment.

Apparatus.—A two platform, all black, elevated runway previously described (Birch, Ison, & Sperling, 1960) was employed. The two platforms are approximately 3 ft. above floor level and enclosed in separate homogeneous black chambers, independently illuminated. A 10-w. bulb mounted on the wall of the approach chamber 3½ ft. above the terminal end of the approach platform provides dim light for this chamber and prevents stray light from entering the goal chamber. In crossing the approach platform (12 in. long and 3 in. wide), *S* breaks a photoelectric beam which turns on overhead illumination of 5.29 ft-L in the goal chamber. The beam is located 1 in. from the end of the approach platform so that the goal platform with its accompanying gap is not visible to *S* until *S* is in close proximity to it. A manually operated guillotine door prevented retracing. The latency of response was measured by a Standard Electric clock from the onset of the goal chamber illumination until *S* removed its weight from the approach platform.

The gap manipulation was imposed by placing a goal platform, 18 in. long and 3½ in. wide, 1 in. from the approach platform for the small gap condition and by placing a platform, 15 in. long and 3½ in. wide, 4 in. from the approach platform for the large gap condition. Thus, for both the small and large gap the terminal end of the goal platform was 19 in. from the approach platform. In addition, the goal platform was 1¼ in. below the height of the approach platform in both gap conditions. A ¼ tsp. aluminum measuring spoon, attached to the end of the goal platform and out of the sight of *S* from the approach platform, served as the goal cup.

Procedure.—All *S*s received extensive handling and pretraining under food deprivation with food reward prior to the experiment. An elevated runway routinely used for pretraining in the laboratory was employed. The *S*s were then placed on a 22½ hr. water deprivation schedule with ad libitum drinking for 1½ hr. for six days and trained on the last four days to drink from the goal cup attached to a grey platform located in the goal chamber of the apparatus. Four *S*s were discarded for failure to drink within 5 min. on any one of four test trials given on Day 4.

---

[1] The second author participated in Exp. I while holding an NSF Cooperative Fellowship.

[2] Experiment II is taken from a senior honors thesis prepared by the third author under the direction of the first author.

During the experiment *S* received six trials a day with the two sequences of small and large gap conditions, SLLSLS and LSSLSL, alternated from day to day. Every trial terminated with *S* drinking 3 drops of water, and the intertrial interval was approximately 20 min. Training continued for 18 days. One group of 13 randomly selected *S*s was trained under 20 hr. deprivation and a second group under 4 hr. deprivation. Food was available at all times in the living cages.

Results

Acquisition data for the four combined conditions of T_D and G over the 18 days of training are shown in Fig. 1 and 2. Because of noticeably slower latencies on the initial trials of each day, only the last four trials of a day are considered. Two latency transformations, reciprocal latency, or response speed (RS), and log latency (LL), were employed. Figure 1 presents the mean RS for the two trials appropriate to each condition on each day while Fig. 2 presents the corresponding mean LL data.

The two figures show effects attributable to training, deprivation, and gap for both transformations. In addition, for RS the gap difference is close to but slightly smaller for 20 hr. thirst than for 4 hr., and for LL the gap difference is considerably smaller for 20 hr. than for 4 hr. The 18 days of training were grouped into three periods of six days for purposes of statistical analysis.

Table 1 summarizes the analyses of variance for RS and LL. Both transformations show all three main effects to be highly significant. Of major interest, however, is the $G \times T_D$ interaction. This interaction is non-

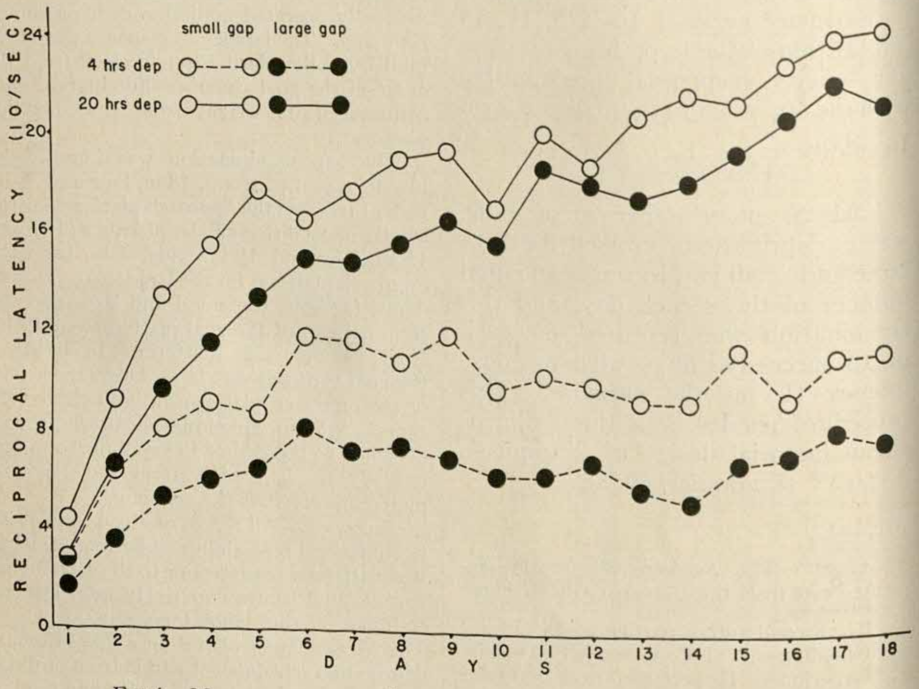


FIG 1. Mean response speed as a function of the size of gap and the degree of water deprivation.

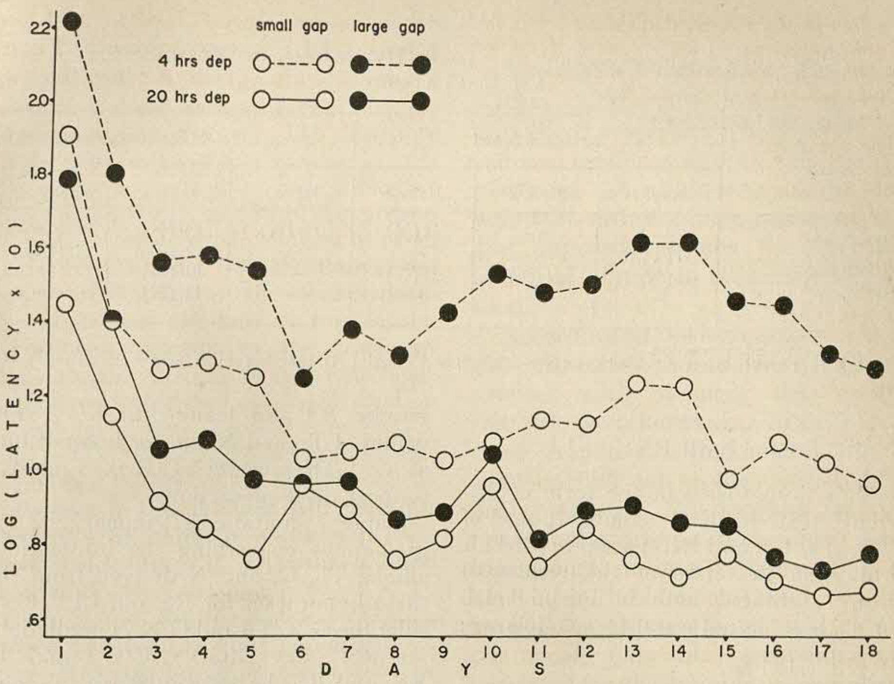


FIG. 2. Mean log latency as a function of the size of gap and the degree of water deprivation.

significant for RS ($P > .20$, $F = 1.68$, $df = 1/24$) but highly significant for LL ($P < .001$, $F = 15.08$, $df = 1/24$). In addition, the $T_D \times N$ interaction is nonsignificant for LL ($P > .10$, $F = 1.78$, $df = 2/48$) but highly significant for RS ($P < .001$, $F = 9.02$, $df = 2/48$), and the $G \times N$ and

TABLE 1

ANALYSES OF VARIANCE OF RESPONSE SPEEDS AND LOG LATENCIES

Source	*df*	Response Speed		Log Latency	
		MS	*F*	*MS*	*F*
Between *S*s	25				
Deprivation (T_D)	1	2972.0	10.10**	7.6711	15.94***
Error (b)	24	294.1		.4813	
Within *S*s	130				
Gap (G)	1	342.0	92.68***	1.9071	74.30***
Training (N)	2	454.4	20.19***	1.1639	19.22***
$G \times N$	2	0.4	<1	.0043	<1
$G \times T_D$	1	6.2	1.68	.3870	15.08***
$T_D \times N$	2	203.1	9.02***	.1078	1.78
$G \times T_D \times N$	2	3.2	2.33	.0223	1.86
Error (w)	120				
E_1	24	3.7		.0257	
E_2	48	22.5		.0606	
E_3	48	1.4		.0120	

** $P = .01$.
*** $P = .001$.

TABLE 2

SUMMARY OF PREDICTIONS FROM THE THREE HYPOTHESES FOR RESPONSE SPEED AND LOG LATENCY

Source	Response Speed			Log Latency		
	Hyp. 1	Hyp. 2	Hyp. 3	Hyp. 1	Hyp. 2	Hyp. 3
$G \times N$	—[a]	—	If(N)	Df(N)	Df(N)	—
$G \times T_D$	If(T_D)[b]	—	—	—	Df(T_D)	Df(T_D)
$T_D \times N$	If(N)	If(N)	If(N)	—	Df(N)	—
$G \times T_D \times N$	—	—	—	—	Df(N)	—

[a] —indicates no interaction.
[b] Effect of first factor is indicated as an increasing, I, or decreasing, D, function of increases in the second factor.

$G \times N \times T_D$ interactions are nonsignificant for both RS and LL.

Three hypotheses in the form of different mathematical combinations of $E(T_D)$, $E(G)$, and $E(N)$ are proposed for evaluation in terms of the results of Exp. I. Stated both in RS and LL form these hypotheses are as follows:

Hypothesis 1 is:

$$RS = E(T_D)[E(N) - E(G)]$$

and

$$LL = -\log E(T_D) - \log[E(N) - E(G)]$$

Hypothesis 2 is:

$$RS = E(T_D)E(N) - E(G)$$

and

$$LL = -\log[E(T_D)E(N) - E(G)]$$

and Hypothesis 3 is:

$$RS = E(N)[E(T_D) - E(G)]$$

and

$$LL = -\log E(N) - \log[E(T_D) - E(G)]$$

A reciprocal relation between response latency and the effects of experimental manipulations, implied by all three hypotheses, has a precedent in the assumptions of Hull (1943) and the derivations of Estes (1950) and Spence (1954), at least for $E(N)$.

Assuming $E(T_D)$, $E(G)$, and $E(N)$ to be some unspecified monotonically increasing functions of T_D, G, and N, respectively, all three hypotheses predict greater RS and lesser LL for greater values of T_D and N and for lesser values of G. The results of Exp. I clearly support these predictions.

Table 2 contains a summary of the predictions concerning the interactions among T_D, G, and N derived from the three hypotheses for RS and LL. Consider the $G \times T_D$ and $G \times N$ predictions for RS. Let $E(G_1) < E(G_2)$ and let RS_1 and RS_2 be associated with G_1 and G_2. Then, for Hyp. 1, $(RS_1 - RS_2) = E(T_D)[E(G_2) - E(G_1)]$; for Hyp. 2, $(RS_1 - RS_2) = [E(G_2) - E(G_1)]$; and for Hyp. 3, $(RS_1 - RS_2) = E(N)[E(G_2) - E(G_1)]$. Thus, for RS Hyp. 1 predicts a $G \times T_D$ interaction while Hyp. 2 and 3 predict no such interaction. In addition, Hyp. 1 and 2 predict no $G \times N$ interaction while Hyp. 3 predicts such an interaction. The directions of the interaction effects are that greater T_D leads to a larger gap difference effect under Hyp. 1 and that greater N gives a larger gap difference effect under Hyp. 3.

Now consider $(LL_2 - LL_1)$ for the three hypotheses. Hypothesis 1 yields:

$$(LL_2 - LL_1) = \log\left[\frac{E(N) - E(G_1)}{E(N) - E(G_2)}\right]$$

Hyp. 2 results in:

$$(LL_2 - LL_1) = \log\left[\frac{E(T_D)E(N) - E(G_1)}{E(T_D)E(N) - E(G_2)}\right]$$

and Hyp. 3 yields:

$$(LL_2 - LL_1) = \log\left[\frac{E(T_D) - E(G_1)}{E(T_D) - E(G_2)}\right]$$

Thus, for LL Hyp. 1 predicts no $G \times T_D$ interaction while Hyp. 2 and 3 predict an interaction, and Hyp 1 and 2 predict a $G \times N$ interaction while Hyp. 3 predicts no interaction. The directions of the interaction effects are that greater T_D leads to a smaller gap difference effect under Hyp. 2 and 3 and that greater N yields a smaller gap difference effect under Hyp. 1 and 2. Similar derivations can be carried out for the three hypotheses under the two latency transformations with respect to the predicted $T_D \times N$ and $G \times T_D \times N$ interactions.

The results for the $G \times T_D$ interaction clearly favor Hyp. 2 and 3 over Hyp 1. Not only is the $G \times T_D$ interaction nonsignificant for RS but the direction of the difference in the obtained gap effect for 20 and 4 hr. thirst is opposite to that predicted by Hyp. 1. Hypotheses 2 and 3 predict the nonsignificant $G \times T_D$ interaction for RS, the significant $G \times T_D$ interaction for LL, and the smaller LL gap differences for 20 hr. thirst than for 4 hr.

A decision between Hyp. 2 and Hyp. 3 is more difficult. The overall pattern of results in comparison with the predictions appears to favor Hyp. 3. Most damaging to Hyp. 2 is the failure to obtain significance for the $G \times N$ and the $T_D \times N$ interactions with LL as predicted. The obtained $G \times N$ effect does agree in direction with Hyp. 2, but the obtained $T_D \times N$ effect does not. The predictions from Hyp. 3 disagree with the data only in the $G \times N$ interaction for RS where the direction of differences is correct but nonsignificant.

The results of Exp. 1 strongly suggest that the effect of a gap on response speed is independent of the degree of thirst, at least for the conditions employed in the experiment. Whether the gap effect changes with training, as implied by Hyp. 3, or is independent of training, as implied by Hyp. 2, is less clearly answered by the experiment. Exp. II was designed to obtain replication data on this point and also to inquire about possible stimulus properties of the gap.

EXPERIMENT II

The major hypothesis under examination in this experiment is that the gap elicits responses which result in a reduced tendency to approach the goal platform. A straight, elevated runway with either a large gap or a small gap imposed between the approach and goal platforms on each trial was used.

According to the hypothesis, a large gap stimulus should result in responses with stronger decremental effects than a small gap stimulus and these responses should tend to become conditioned to the stimuli of the goal platform. By making the approach platform and the goal platform alike in one brightness (e.g., black) for the large gap and alike in another brightness (e.g., white) for the small gap, an opportunity is provided for the two brightness stimuli to become conditioned differentially to the two sets of interfering responses and to elicit these responses differentially on the approach platform. Thus, a result consistent with the hypothesis would be that, as training continues, the response time on the approach platform prior to the small gap drops below the corresponding time prior to the large gap, even though the gap stimuli themselves are not visible in this section of the runway.

The experiment also includes replication data relevant to a decision between the adequacy of Hyp. 2 and Hyp. 3 as presented in Exp. I.

Method

Subjects.—The *S*s were 16 experimentally naive female pigmented rats from the same colony as those in Exp I, approximately 120 days old at the beginning of the experiment. Two *S*s were discarded for failure to run on the second day of training.

Apparatus.—An elevated runway similar to that of Exp. I was used. Three platforms,

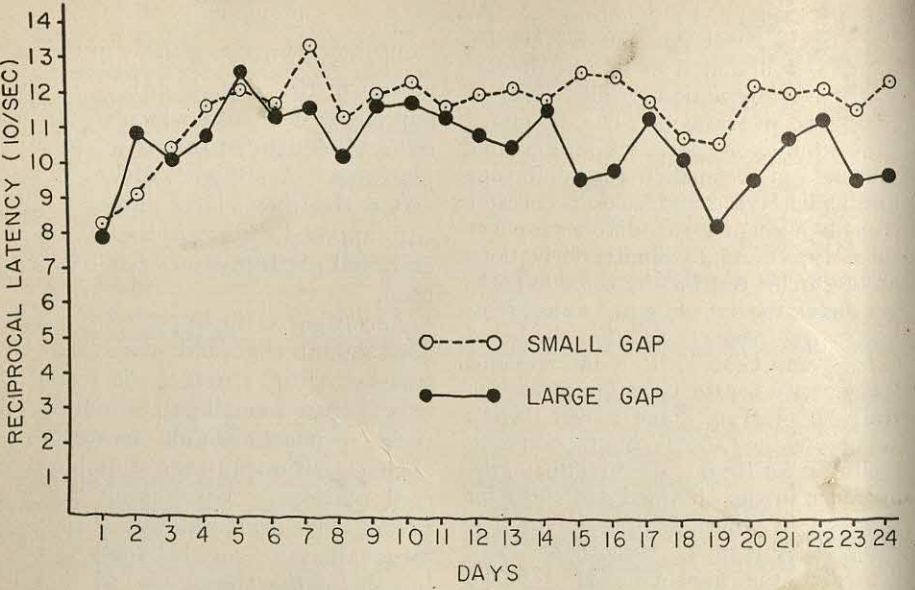


FIG. 3. Mean response speed on approach runway for large and small gap conditions.

start, approach, and goal, were housed in separate chambers to permit control of the presentation of the platform stimuli to *S*. The dimensions of the start, approach, and goal platforms were 3 in. wide by 12 in., 24 in., and 18 in. long, respectively.

The arrangement of photoelectric beams and microswitches under the platforms provided measures of the time taken by *S* to traverse the approach platform and to cross from the approach to the goal platform.

TABLE 3

ANALYSIS OF VARIANCE OF RESPONSE SPEEDS ON APPROACH PLATFORM

Source	*df*	*MS*	*F*
Between *S*s	13		
Groups (Gr)	1	74.7	3.14
Error (b)	12	23.8	
Within *S*s	154		
Gap (G)	1	43.4	22.77***
Training (N)	5	11.8	1.86
G × N	5	3.9	4.36**
G × Gr	1	5.0	2.60
Gr × N	5	2.2	<1
G × Gr × N	5	0.4	<1
Error (w)	132		
E_1	12	1.9	
E_2	60	6.3	
E_3	60	0.9	

** $P = .01$.
*** $P = .001$.

Raising a guillotine door permitted *S* to leave the start platform and also activated three Standard Electric clocks. The first clock stopped when *S* placed its weight on the approach platform, the second clock stopped when *S* broke a photoelectric beam 1 in. from the end of the approach platform, and the third clock stopped when *S* placed its weight on the goal platform. Breaking the photoelectric beam also illuminated the goal chamber. The approach platform latency was given by the difference in readings from the first and second clocks, and the gap crossing latency by the difference from the second and third clocks.

The *S* stepped down 1 in. from the start to the approach platform and another 1 in. from the approach to the goal platform. Under the small gap condition the goal platform was adjacent to the approach platform while the two platforms were separated by 3 in. for the large gap condition. Two sets, black and white, of start, approach, and goal platforms were used.

Procedure.—The *S*s were placed on a 22 hr. food deprivation schedule and allowed to eat for 1 hr. for 22 days prior to the experiment. Preliminary training consisted of 2 days handling, followed by 14 days running on the pretraining elevated runway with food available and 6 days of placing and eating on the goal platforms of the apparatus.

During the experiment *S*s received six trials a day with four sequences of small and large gap conditions, SLSSLL, SSLLSL, LLSSLS, and LSLLSS, repeated over the 24 days of training. For Group A ($N = 8$) all three platforms were black under the large gap condition and white under the small gap condition. Group B ($N = 6$) received the opposite pairing of the platform brightnesses with the gap conditions. Each trial terminated with *S* eating two small pellets of laboratory food (.1 gm.) from a recessed food cup at the end of the goal platform. The intertrial interval was approximately 8 min.

TABLE 4

Mean Response Speed and Mean Log Latency of Gap Crossing for Groups, Gap Conditions, and Training

Group	Gap	Training					
		I	II	III	IV	V	VI
Response Speed (Reciprocal Latency × 10)							
A	Small	5.1	10.6	15.2	16.5	13.4	12.5
	Large	1.6	4.4	5.3	5.2	4.0	4.8
B	Small	8.2	22.2	18.7	20.1	13.2	16.5
	Large	4.2	12.9	13.2	12.4	9.9	11.2
Log (Latency × 10)							
A	Small	1.72	1.32	1.11	1.20	1.38	1.25
	Large	2.11	1.68	1.55	1.66	1.78	1.68
B	Small	1.42	.84	.98	1.03	1.07	.97
	Large	1.73	1.07	1.20	1.26	1.33	1.30

Results

Figure 3 shows the mean response speed (reciprocal latency × 10) on the approach runway for Groups A and B combined under the small and large gap conditions over the 24 days of training. The middle four trials of each day were used to compute all mean values. These data indicate that response speed is not different for the two gap conditions at the beginning of training but becomes relatively faster under the small gap condition than under the large gap condition as training continues. Thus, over the period of training the differential pairing of the black and white platforms with the two gaps produces different response speeds on the approach platform prior to the actual presentation of the gap condition.

A summary of the analysis of variance on these data is presented in Table 3. Of particular interest are the significant values obtained for the Gap main effect ($P < .001$) and

TABLE 5

Analyses of Variance of Response Speeds and Log Latencies of Gap Crossing

Source	*df*	Response Speed		Log Latency	
		MS	*F*	*MS*	*F*
Between *S*s	13				
Groups (Gr)	1	1184.4	4.58	5.1244	6.42*
Error (b)	12	258.4		.7988	
Within *S*s	154				
Gap (G)	1	2100.2	48.12***	5.1345	188.77***
Training (N)	5	273.1	4.72**	1.0941	6.47***
G × N	5	27.6	1.64	.0043	<1
G × Gr	1	46.1	1.06	.2242	8.24*
N × Gr	5	47.5	<1	.0784	<1
G × N × Gr	5	19.9	1.18	.0057	<1
Error (w)	132				
E_1	12	43.6		.0272	
E_2	60	57.9		.1690	
E_3	60	16.8		.0183	

* $P = .05$.
** $P = .01$.
*** $P = .001$.

$G \times N$ interaction ($P < .01$), both findings being consistent with the hypothesis of Exp. II.

Mean response speed and mean log latency over blocks of 4 days for gap crossing for Groups A and B are given in Table 4. Table 5 presents the summary of the analyses of variance on these data. The main effect for gap is highly significant both for RS ($P < .001$), and for LL ($P < .001$), but the $G \times N$ interaction fails to reach significance under either latency transformation. Examination of Table 4, however, shows a trend for greater gap effect differences in RS with increased training similar to that found in Exp. I.

Discussion

The replication data of Exp. II on the $G \times N$ interaction are inconclusive. The finding of both experiments is that no significant $G \times N$ effect was obtained for RS or LL, but in each there is the suggestion that the gap differences may increase with training for RS, as predicted by Hyp. 3. In view of the small *F* values obtained for the $G \times N$ interaction in the two experiments, supporting the lack of appearance of any strong effect in Fig. 1, it would appear more desirable to collect additional data oriented specifically to this problem rather than exploit the present data with further analyses.

A final caution in the interpretation of the three hypotheses is warranted. A more complete mathematical statement of the latency of response should include the time required to carry out the response, given that it is elicited (Spence, 1954). The present three hypotheses are not intended to include this latency component and are adequately tested in the present experiments only under the assumption that its effect is negligible. This assumption is more clearly justified in Exp. I where the gap crossing latency was measured from the onset of the goal chamber illumination to *S*s leaving the approach platform rather than to *S*s reaching the goal platform as in Exp. II.

Summary

Two experiments obtained data relevant to the theoretical interpretation of the effect of size of gap on the response latency of rats in a straight elevated runway. In Exp. I, time of water deprivation (T_D), size of gap (G), and amount of training (N) were combined factorially, and the form of the mathematical expression relating the effects of these manipulations, $E(T_D)$, $E(G)$, and $E(N)$, to response speed (RS), the reciprocal of the response latency, was examined. Three mathematical combinations of $E(T_D)$, $E(G)$, and $E(N)$ were proposed and tested by using both a reciprocal and a logarithmic transformation on the latencies. Hypothesis 1, $RS = E(T_D)[E(N) - E(G)]$, was found inadequate and the total pattern of results seemed to favor Hyp. 3, $RS = E(N)[E(T_D) - E(G)]$, over Hyp. 2, $RS = E(T_D)E(N) - E(G)$.

In Exp. II it was hypothesized that the gap elicits responses that produce a decremental effect on the tendency to approach the goal platform. Under this hypothesis it was expected that two degrees of interfering responses arising from the large and small gaps could be conditioned differentially to two sets of stimuli and that these two degrees of interference could be exhibited in the absence of the gap stimuli themselves. The results supported the hypothesis in that, as training continued, the response time on the approach platform prior to the small gap fell below the corresponding time prior to the large gap, even though the gap stimuli themselves were not visible in this section of the runway.

References

Birch, D., Ison, J. R., & Sperling, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Spence, K. W. The relation of response latency and speed to the intervening variables and *N* in S-R theory. *Psychol. Rev.*, 1954, **61**, 209–216.

(Received June 14, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 187–195

THE RELATIONSHIP BETWEEN THE GALVANIC SKIN RESPONSE, VASOCONSTRICTION, AND TACTILE SENSITIVITY[1]

ROBERT EDELBERG

Baylor University College of Medicine

In the last decade increasing attention has been given to the screening action or selectivity which the central nervous system exerts over the intake of environmental events (Lindsley, 1956; Livingston, 1959). These investigations have concerned two mechanisms of screening: (*a*) by control of the sensitivity of receptors, and (*b*) by a selective action in the more central processing of the arriving signals either at the sensory nuclei or at higher relay stations such as the thalamus (King, Naquet, & Magoun, 1957). Central control over the threshold of peripheral receptors or primary synapses by tonic bombardment along efferent fibers has been demonstrated in the case of muscle spindles by Granit and Kaada (1952) and by Hagbarth and Kerr (1954), in the case of the vestibular system by Gernandt and Thulin (1955), the retina by Granit (1955), olfactory receptors by Kerr and Hagbarth (1955) and by Tucker and Beidler (1956), and tactile receptors by Loewenstein (1956b). Another mechanism for control of peripheral reception is exemplified in the contraction of the muscles of the middle ear, with an ensuing reduction of auditory sensitivity either in response to loud sounds (Galambos & Allen, 1958) or following reticular stimulation (Hugelin, Dumont, & Paillas, 1960).

A related group of studies deals with the screening mechanism on an integrated behavioral basis. Demonstration of selectivity as a consequence of focus of attention is described in the oft-cited experiment of Hernandez-Peon and his co-workers (1956), recently challenged in part by Hugelin et al. (1960), who showed that attention by an unanaesthetized cat to mice, fish odors, and electric shocks reduced the cochlear nucleus potentials evoked by clicks. Additional evidence for the role of attention in modifying sensory intake is seen in the work of Callaway (Callaway & Dembo, 1958) and supported in an investigation by Lacey (1959). These authors indicate that central activity will tend to regulate the degree to which the organism is receptive to specific aspects of the environment peripheral to central attention in terms of space, time, or meaning (Callaway & Dembo, 1958), or to its environment as a whole (Lacey, 1959). Also in the area of focus of attention is the effect of warning on visual and auditory threshold, the minimum threshold occurring some short interval after a warning signal (Howarth & Treisman, 1958).

It was with regard to the possible control of the sensitivity of receptors that the original investigation of tactile sensitivity was undertaken at

---

[1] This investigation was supported in part by a research grant, M-1904, from the National Institute of Mental Health, United States Public Health Service; in part by Air Force Contract AF 18 (603)-79, Air Force Office of Scientific Research; and in part by a Senior Research Fellowship, SF-219, from the United States Public Health Service. The author is grateful for valuable suggestions by Neil R. Burch and for the competent assistance of George L. Johnson.

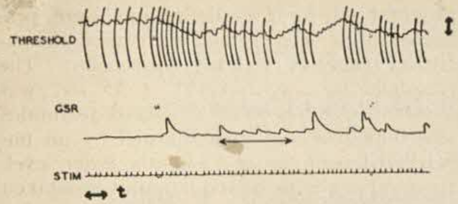


FIG. 1. Recording of tactile threshold and GSR. (Upper trace: continuous line indicates intensity of vibration; vertical lines indicate when *S* perceived stimulus. Stimulus trace: vertical lines indicate 1-sec. stimulus delivered to *S*. Time line, t, 10 sec. Calibration line at end of threshold trace indicates a change of 2 db.)

this laboratory. Darrow (1933) had suggested that the galvanic skin response (GSR) is a means by which the organism facilitates grasping by the secretion of sweat which moistens the skin and thereby mechanically increases the tactile sensitivity. It was thought by the present author that if tactile sensitivity were in fact related to GSR activity, the mechanism might be a direct sensitizing of peripheral receptors by sympathetic efferents rather than a mechanical consequence of sweating. Preliminary experiments (Edelberg & Burch, 1958) indicated first, that there was clearly a relationship between GSR and tactile sensitivity, the threshold usually falling as GSR activity increased, and secondly, contrary to Darrow, that moisture played no part in this phenomenon since the results were as easily demonstrated when the skin was under water as when it was dry. These first experiments had been carried out under conditions in which large GSRs were elicited after a baseline of low autonomic activity had been reached. When conditions were altered to allow greater baseline activation of the subject, the correlation, though manifest, became considerably less impressive, and doubt was cast upon the initial interpretation that the GSR might be a by-product of the arrival of nerve impulses serving to sensitize tactile receptors. Rather, the GSR may simply have been reflecting changes in central arousal which in turn influence tactile sensitivity. Further experiments have been conducted to help resolve this question and to extend the observations to cover concomitant vasomotive activity; the results are reported in this paper.

METHOD

Apparatus

Electrode application, masking of a fixed area, use of a constant current source, and GSR recording equipment were similar to those reported elsewhere (Edelberg, Greiner, & Burch, 1960) except that K-Y surgical jelly was used as the contact medium. The plethysmograph consisted of a water-filled chamber maintained at 30°C. coupled to a volume-sensitive electrical transducer having a full-scale sensitivity of 40 mm^3.

Weitz (1939) and Geldard (1940) have presented evidence to show that the responses to vibration and gentle pressure are mediated by the same sensory system. Consequently, for reasons of convenience, 1-sec. transientless pulses of 250-cps vibration were presented as tactile stimuli. These were produced by a phonograph cutting head fitted with a blunt probe and resting on the palmar skin of the distal segment of the index finger with a constant force of 1 gm., this being achieved by a counterbalanced lever and pivot arrangement. The end of the probe was spherical and was 2.6 mm. in diameter. The variable ac voltage from the oscillator was monitored, rectified, and recorded. The stimulus level was thus written out as a continuous curve (Fig. 1), but only at the instants marked by the signal on the lower trace was an automatically timed pulse delivered to the cutting head. All switching operations were noiseless. The calibration was accomplished with a General Radio Type 1551-A sound level meter fitted with a General Radio vibration pickup, Type 759-P35. Vibration intensity could be varied from 25 to 40 db. (re .0002 microbar), the zero current background vibration being 25 db. In most cases the threshold was around 30 db., varying between 26 and

36 db. A range switch allowed selection of several intensity ranges, each of which started at 25 db. (RMS amplitude = .45 μ), i.e., with zero current to the cutting head.

Subjects and Procedure

The *S*s were 28 paid volunteers of both sexes with age range from 18 to 36 yr., all naive except 2 (identified in Table 1). The *S* was told that the experiment was for the purpose of measuring touch sensitivity, and that the electrodes and plethysmographic chamber were for studying other measures at the same time. Gain was adjusted so that the GSR in response to a rapid maximal deep breath approximated full scale deflection. Tactile stimuli without cue were presented at approximately 2-sec. intervals. If *S* perceived the vibration, he depressed the foot switch which recorded a vertical line on the stimulus trace and which also signaled *E* who sent the next pulse at a lower intensity. If the stimulus was not perceived, *E* increased the intensity before delivering the next pulse. After a training and equilibration period which usually took about 20 min., the actual tactile testing would procede uninterruptedly for a minimum of 5 min. or until a 5-min. sequence at the same stimulus range setting could be obtained. A 250-cps masking tone from a loud-speaker insured that *S* could not hear any vibration from the cutting head probe. At random intervals, *S* was startled by a noise or was asked to sniff sharply or to take a deep breath. Not more than 3 autonomic responses were thus deliberately elicited during the 5-min. period. Autonomic activity was otherwise spontaneous, no special effort being made to wait for fullest relaxation and minimum GSR activity as in the earlier series. The external stimulus, however, was maintained at a low level with regard to noises and movements.

Treatment of the Data

GSR amplitudes were measured in terms of resistance change from onset to peak of the wave, waves less than $\frac{1}{10}$ full scale (see procedure) being disregarded. Two measures of vasomotor activity were used: *pulse volume* (volume change from trough to peak of a pulse wave) and *relative volume* (absolute volume minus an unmeasured constant). Averages were obtained for 15-sec. periods by a graphical method of interpolation.

Tactile thresholds were determined separately for appropriate portions of a 15-sec. period and the results averaged on a time-prorated basis. Pen deflections were proportional to decibels and were used in statistical analyses without conversion. The threshold for any interval of 15 sec. was determined as follows: If a perceived stimulus was both preceded and followed by an unperceived stimulus at a slightly lower level, the level of the perceived stimulus was taken as the threshold. If a perceived stimulus was preceded by an unperceived stimulus, but followed by two or more perceived stimuli of decreasing intensity (e.g., first major fall in threshold, Fig. 1) it was assumed that the method of search may not have allowed for determination of the new threshold in as short a time as it was actually lowered. Tests in which *E* decreased intensity in large steps showed this assumption to be valid. Consequently, the lowest perceived stimulus in this train within the 15-sec. period was taken as the level back to the first perceived stimulus. On a similar experimental basis, if a perceived stimulus was followed by an unperceived stimulus, and this unperceived stimulus followed by a train of unperceived stimuli of increasing intensity, it was assumed that the threshold for this period of no response was the level of the first perceived stimulus following this train. If this 15-sec. period ended prior to the perception of a stimulus, the highest stimulus of the train within the period was taken as threshold back to the first unperceived stimulus of the train.

Although all data were written out on the same record, the analysis of each variable was accomplished by separate *O*s with the remaining variables covered up during analysis.

Statistical treatment.—A graph of threshold as a function of relative volume or pulse volume indicated the linearity and distribution of the data to warrant application of a product-moment analysis for correlations. Such did not hold for GSR, and in this instance a chi square test was used for determining concordance between falls in tactile threshold and incidence of GSRs. When spontaneous GSR activity was low, an elicited GSR was almost always followed by a significant fall in tactile threshold. The association of these events, each appearing conspicuously against a stable background, was clearly apparent, but the lag varied from 1 to 10 sec. As discussed elsewhere, distraction of *S* by the arousing stimulus could easily cause a lag of this order before the lowered threshold became manifest. Consequently, if a GSR was followed within 10 sec. after onset by a fall in tactile threshold, the case was counted as a simultaneous occurrence. If the tactile fall preceded the GSR, the events were not considered simultaneous since inspection of the temporally isolated cases showed the GSR to precede the tactile change. The time not taken up by GSRs and/or fall in tactile threshold was divided into 10-sec. intervals to complete the 2 × 2 table. A similar technique was used for determining the chi square relationship between decrease in pulse volume and fall in threshold.

RESULTS

GSR.—The correlation of tactile sensitivity with GSR depended in part upon state of alertness. An *S* who showed a GSR record which was, in fact, a continuous series of waves, showed little relationship between larger responses and tactile sensitivity. On the other hand, in *S*s in a relaxed state with little activity, a sudden onset of GSR activity was ordinarily associated with a significant fall (as much as 4 db.) in threshold to tactile stimulation. Note that the first large GSR of Fig. 1 is associated with a conspicuous fall in threshold while the group of GSRs above the central arrow is associated more with a constant low threshold than with rises and falls.

On the whole, the 24 *S*s of this series showed a low correlation as compared with the 8 of the previous study who were tested under conditions designed for evoking large GSRs against a background of low level GSR activity. A chi square analysis of the pooled data for the entire earlier group showed an association ($\phi = .71; P < .01$) between GSR and fall in threshold. In the present series, the pooled data on 24 *S*s showed only a .43 (phi) association ($N = 703; P < .01$). For the 12 *S*s who had both plethysmograms and GSRs recorded, the GSR-tactile association was .41 ($N = 346; P < .01$). When the data for the 24 *S*s were analyzed with the exclusion of all periods of 30 sec. or longer showing a GSR frequency greater than 4 per min., ϕ rose to .52 ($N = 442; P < .01$).

One factor tending to obscure the tactile-autonomic relationship was variation in attention. The intervention of this factor was demonstrated by a simple experiment. A series of stimuli was given at a constant level of intensity, slightly suprathreshold, until *S* ceased responding for approximately five signals. At this point one stimulus was delivered at a somewhat higher level (e.g., by 1 db.) and the next stimulus given at a level slightly lower than the previous series. Invariably this maneuver resulted in a recovery of *S*'s response even though no signs of arousal either in the plethysmogram or in skin resistance were noted. The *S*s reported that their attention wandered but that feeling a vibration would frequently bring them back to their task. The return of attention to the tactile task after temporary diversion is in fact considered a probable source of many of the falls in threshold which occurred without accompanying autonomic activity. On the other hand, autonomic activity was frequently not accompanied by a fall in threshold, a situation to be expected if *S*'s attention had drifted from the task at this time. The attention factor also worked against a correlation in many cases when *S* was startled in an effort to induce autonomic activity. In some *S*s this startle, although associated with a high autonomic response, was followed consistently by a rise in threshold and a little later on by a drop to lower than the previous level.

Vascular responses.—Finger pletheysmograms were obtained from 12 of the above *S*s simultaneously with the GSR and tactile threshold determination (Table 1 and Fig. 2). A recording of an exceptionally good relationship is seen in Fig. 3. This record illustrates a case in which a threshold change is associated with vasomotive activity but not GSR (see section of record above arrow); in other sections of the same record both GSRs and vasoconstriction occur in the same time period and are associated with a fall in threshold.

The product-moment correlation between pulse volume and tactile threshold was greater than .56 ($P \leqq .01$) in 5 out of the 12 cases, and an additional 3 *S*s show $r > .44$ ($P \leqq .05$). For relative volume and tactile threshold, 4 of the 12 showed a correlation greater than .57 ($P \leqq .01$). The chi square comparison for tactile threshold and pulse volume changes for the pooled data of all 12 *S*s (10-sec. samples) showed $\phi = .51$ ($P < .01$).

TABLE 1

Product-Moment Correlations between Threshold and Plethysmogram Data for Each of 12 *S*s

S	*N*	Threshold vs. Pulse Volume		Threshold vs. Volume	
		r	$P \leqq$	*r*	$P \leqq$
1[a]	20	.50	.05	.68	.01
2	20	.44	.05	.46	.05
3	20	.60	.01	.59	.01
4	20	.35		−.13	
5[a]	20	.48	.05	.32	
6	16	.63	.01	.32	
7	18	.67	.01	.22	
8	20	−.35		−.19	
9	20	.62	.01	.57	.01
10	19	−.21		−.18	
11	18	.56	.01	.72	.01
12	20	−.31		−.45	.05

[a] Sophisticated *S*.

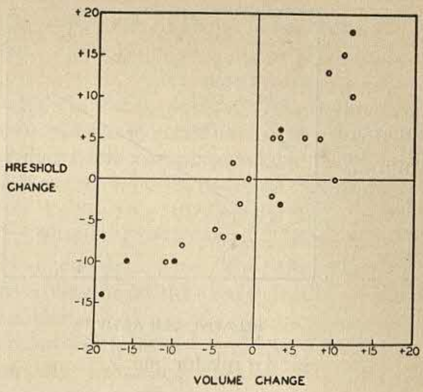


Fig. 2. Relationship between change in tactile threshold and change in finger volume for a single *S*. (Units are arbitrary. Solid circles indicate simultaneous GSR occurred.)

Mechanical effects of vasomotion.—Since finger volume changes alter the stretch on the skin, a study of the mechanical effect of such stretching was undertaken. Stretching accomplished by applying masking tape to the skin increased the threshold by 2 to 4 db. Engorgement of the finger by inflation of an arm cuff to 60 mm. Hg produced a volume change of the order of that occurring during natural vasomotive activity and coincidently elevated threshold by about 1 db. Both effects were reproducible and readily reversible.

Anatomical correlations.—The work of Loewenstein (1956b) on frog skin and on the Pacinian corpuscle of the cat (Loewenstein & Altamirano-Orrego, 1956) pointed to a possible mechanism for tactile threshold changes involving sensitization of peripheral receptors by activity of sympathetic efferents. On the assumption that

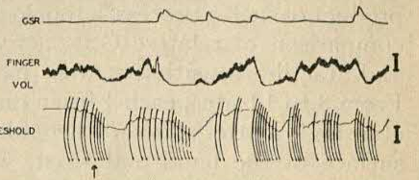
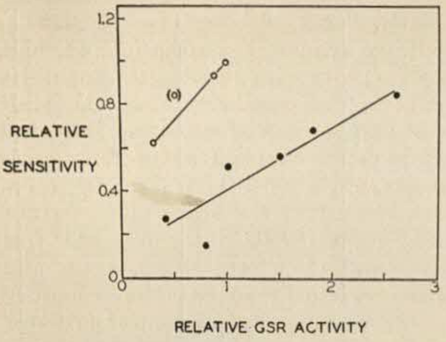
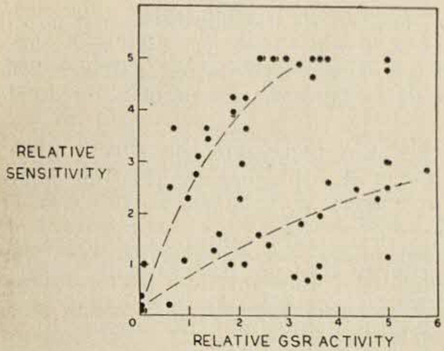


Fig. 3. Recording of GSR, plethysmogram and tactile threshold for a single *S*. (Time line, t, 10 sec. Calibration lines at end of record: volume change, 10 mm.^{3}; threshold change, 1 db. The three large signals on the stimulus trace—bottom—indicate *S* sniffed.)



4a. Graph for one *S*.



4b. Graph for 6 *S*s showing two populations of sites.

FIG. 4. Relation between tactile sensitivity and relative GSR amplitude for various sites scattered over palm. (In Figure 4a, open circles from points along lateral divisions of median nerve; solid circles from along medial division of median nerve. Point in parentheses lay between the two branches and could not be definitely assigned to either division.)

the GSR might be in part a by-product of such a process, a topological comparison of relative GSR activity and tactile sensitivity was made. From 8 to 12 sites, each $\frac{1}{4}$ in. in diameter and scattered over the palmar surface of the hand and wrist, were delineated by 1-in. masking tape squares having a $\frac{1}{4}$-in. hole punched in them. Since the variation of GSR activity with arousal (Burch & Greiner, 1958) makes sequential comparisons unreliable, the method of 2-site simultaneous recording was used as described elsewhere (Edelberg et al., 1960). The relative activity of each site was obtained by comparing it separately with the simultaneously observed reference site and expressing the results as ratios. The same approach was used for tactile sensitivity determinations, one site being maintained as a fixed tactile reference, while a second tactile stimulator was used to test the sites to be compared. Since simultaneous tactile testing of both sites could not be accomplished, three sets of determinations on the reference site and on the variable site were made alternately over a relatively short, quiet period (approximately 2 min.) and the results averaged. After completion of the tactile series a drop of electrode paste (K-Y jelly) was placed on each site in turn, an electrode applied, and three GSRs evoked by having *S* sniff or take a deep breath. The reference site was prestabilized and the exposure time of the experimental sites to the K-Y jelly held constant.

Results of topological survey.—Plots of the relationship between tactile sensitivity and relative GSR intensity for each site showed the sites falling into two populations depending on proximity to either the ulnar or median nerve and even on proximity to either the lateral or medial division of the median nerve. The results of one such experiment are shown in Fig. 4a and the results for all 6 *S*s in Fig. 4b. The values for each *S*, independent of site location, have been divided by a constant ($\frac{1}{5}$ the value of *S*'s highest sensitivity) in order to facilitate comparison on a single graph. Despite a wide scatter, the occurrence of two populations of sites, one of high tactile sensitivity, the other low, and the relationship of the relative GSR activity of an area to its relative tactile sensitivity is clearly demonstrated. The demonstration of areas of high and low sensitivity associated with different nerves is in agreement with the recent report of Uttal (1960) on electrical stimuli.

DISCUSSION

Experimental evidence can be cited in support of three possible reasons for the short term relationship between tactile sensitivity and autonomic activity: (*a*) peripheral sensitization due to a mechanical effect; (*b*) peripheral sensitization due to sympathetic impulses; (*c*) central arousal associated concurrently with autonomic activity and increased perceptivity.

Mechanical effect.—With respect to the first possibility, it has been demonstrated here that measures tending to increase the stretch on the skin will also raise the threshold. This observation seems at first glance inconsistent with that of Loewenstein (1956a) who showed that static stretching of frog skin actually causes fast adapting touch receptors to discharge continuously. However, he goes on to say that in some stretched preparations there is "a clear reduction in the number of discharged impulses in response to the (constant level) adequate mechanical stimulus as compared with that discharged at low skin tension" (p. 598), a situation similar to the present investigation. The threshold changes accompanying vasoconstriction could, then, be the consequence of a reduction in the stretch of the skin. However, a fall in threshold was frequently observed without a change in volume. In addition, some *S*s with little vasomotive activity exhibited active threshold variations associated with GSR activity. Lastly, there was a considerable variability in the phase relation between the onset of pulse or volume change and the fall in threshold. These considerations would indicate that even if the mechanical effect is operating, there must be an additional independent mechanism for lowering threshold.

Sympathetic effect.—The spatial as well as temporal association of high GSR activity with high tactile sensitivity suggests the possibility that the GSR is in part a by-product of the arrival of sympathetic impulses whose function is to elevate the sensitivity of tactile receptors located in the epidermis. Arthur and Shelley (1959) have shown that free nerve endings enter the epidermis and have concluded that these have a sensory function. Release of a chemical mediator around such nerve endings could incidentally influence the permeability of surrounding cells and hence alter electrical resistance. This would be consistent with Wilcott's (1959) conclusions that the epidermis as well as the sweat gland constitutes an active site in the GSR. While several arguments may be raised against this hypothesis, in particular the variation in phase relationship and the numerous instances of independent occurrences, it may be defended if other mechanisms for alteration of tactile sensitivity are allotted a dominant role.

Central processes.—There remains the possibility that there is *no* peripheral component to the sensitization process, that autonomic activity and threshold change are unrelated except insofar as they are reflections of the same central arousal process. Silverman, Cohen, and Shmavonian (1959) have favored this viewpoint in their interpretation of the reduction of threshold to electric shock associated with GSR activity. Numerous references to the increase in autonomic activity associated with central arousal might be cited (e.g., Darrow, Jost, Solomon, & Mergener, 1942; Jung, 1954). A type of control at higher centers not necessarily involving increased general arousal would be alterations in attention, as demonstrated neurophysiologically in the experiment of Hernandez-Peon (1956), and as evidenced in the present study. On the basis of experiments reported here, it seems that a central effect must account for much of the demonstrated relationship but that an additional peripheral mechanism has not been ruled out.

Future experiments will involve the testing of *S*s with a history of unilateral sympathectomy to determine whether the sensitizing mechanism has a peripheral component or is entirely central.

In the present stage of the investigation one can safely conclude only this: that there is a mechanism by which the human, while in a low-level external stimulus field, in response to an alerting stimulus or to internal activation, can reduce his tactile threshold during the time of this activation; that this mechanism is prompt and readily reversible upon cessation of the arousal situation; that autonomic activity is normally associated with this sensitization in a close time relationship. A by-product of the investigation is the observation that the magnitude of these shifts in tactile sensitivity is such that any experiment based on the assumption that there is a static tactile threshold in a given area is invalid unless the state of arousal of the organism is maintained constant in some way, or unless changes in arousal are taken into account.

SUMMARY

The relationship over short periods of time between cutaneous tactile threshold and autonomic activity, as reflected in galvanic skin response (GSR) or in degree of vasoconstriction, was determined by a method allowing continuous monitoring of *S*'s threshold to 250 cps vibration. Increased autonomic activity occurring either spontaneously or as a result of external stimulation is generally accompanied by a fall in tactile threshold, the highest correlation occurring when finger pulse volume is used as the autonomic measure. Experiments to determine whether the correlation with vascular activity is due to the mechanical effect of the stretching or relaxing of the skin on tactile threshold showed that this is a possible explanation but cannot account for all of the results.

A survey of the palmar surface of the hand showed a relationship between the relative magnitude of the GSR from a given site and its tactile sensitivity, and further, that the sites comprise separate populations, each associated with a different cutaneous nerve. It is postulated that the GSR may *in part* be a by-product of the arrival of sympathetic impulses whose function is to sensitize tactile receptors. The influence of general arousal and of focus of attention on the tactile-autonomic relationship has tentatively been assigned a dominant position.

REFERENCES

ARTHUR, R. P., & SHELLEY, W. B. The innervation of human epidermis. *J. invest. Derm.* 1959, **32**, 397–411.

BURCH, N. R., & GREINER, T. H. Drugs and human fatigue: GSR parameters. *J. Psychol.*, 1958, **45**, 3–10.

CALLAWAY, E., & DEMBO, D. Narrowed attention: A psychological phenomenon that accompanies a certain physiological change. *AMA Arch. Neurol. Psychiat.*, 1958, **79**, 74–90.

DARROW, C. W. The functional significance of the galvanic skin reflex and perspiration on the backs and palms of the hands. *Psychol. Bull.*, 1933, **30**, 712.

DARROW, C. W., JOST, H., SOLOMON, A. P., & MERGENER, J. C. Autonomic indications of excitatory and homeostatic effects on the electroencephalogram. *J. Psychol.*, 1942, **14**, 115–130.

EDELBERG, R., & BURCH, N. R. Possible sensitizing of tactile receptors by the galvanic skin reflex. *Fed. Proc.*, 1958, **17**, 39.

EDELBERG, R. GREINER, T., & BURCH, N. R. Some membrane properties of the effector in the galvanic skin response. *J. appl. Physiol.*, 1960, **15**, 691–696.

GALAMBOS, R., & ALLEN, R. Action of middle ear muscles in normal cats. *Fed. Proc.*, 1958, **17**, 50.

GELDARD, F. A. The perception of mechanical vibration. *J. gen. Psychol.*, 1940, **22**, 243–308.

GERNANDT, B. E., & THULIN, G. A. Reciprocal effects upon spinal motoneurons from stimulation of bulbar reticular formation. *J. Neurophysiol.*, 1955, **18**, 113–129.

GRANIT, R. Centrifugal and antidromic effects on ganglion cells of retina. *J. Neurophysiol.*, 1955, **18**, 388–411.

GRANIT, R., & KAADA, B. R. Influence of stimulation of central nervous structures on muscle spindles in cat. *Acta physiol., Scand.*, 1952, **27**, 130–160.

HAGBARTH, K.-E., & KERR, D. I. B. Central influences on spinal afferent conduction. *J. Neurophysiol.*, 1954, **17**, 295–307.

HERNANDEZ-PEON, R., SCHERRER, H., & JOUVET, M. Modification of electric activity in cochlear nucleus during attention in unanesthetized cats. *Science*, 1956, **123**, 331–332.

Howarth, C. I., & Treisman, M. The effect of warning interval on thresholds. Flying Personnel Research Committee Report No. 1034, British Air Ministry, 1958.

Hugelin, A., Dumont, S., & Paillas, N. Tympanic muscles and control of auditory input during arousal. *Science*, 1960, **131**, 1371–1372.

Jung, R. Correlation of bioelectrical and autonomic phenomena with alterations of consciousness and arousal in man. In, *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954. Pp. 310–344.

Kerr, D. I. B., & Hagbarth, K.-E. An investigation of olfactory centrifugal fiber system. *J. Neurophysiol.*, 1955, **18**, 362–374.

King, E. E., Naquet, R., & Magoun, H. W. Alterations in somatic afferent transmission through the thalamus by central mechanisms and barbiturates. *J. Pharmacol. exp. Ther.*, 1957, **119**, 48–63.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In, *Research in psychotherapy.* Washington, D. C.: American Psychological Association, 1959. Pp. 160–208.

Lindsley, D. B. Physiological psychology. *Annu. Rev. Psychol.*, 1956, **7**, 323–348.

Livingston, R. B. Central control of receptors and sensory transmission systems. In, *Handbook of physiology.* Washington, D. C.: American Physiological Society, 1959. Pp. 741–760.

Loewenstein, W. R. Excitation and changes in adaptation by stretch of mechanoreceptors. *J. Physiol.*, 1956, **133**, 588–602. (a)

Loewenstein, W. R. Modulation of cutaneous receptors by sympathetic stimulation. *J. Physiol.*, 1956, **132**, 40–60. (b)

Loewenstein, W. R., & Altamirano-Orrego, R. Enhancement of activity in a Pacinian corpuscle by sympathomimetic agents. *Nature, Lond.*, 1956, **178**, 1292–1293.

Silverman, A. J., Cohen, S. I., & Shmavonian, B. M. Investigation of psychophysiologic relationships with skin resistance measures. *J. psychosom. Res.*, 1959, **4**, 65–87.

Tucker, D., & Beidler, L. M. Autonomic nervous system influence on olfactory receptors. *Amer. J. Physiol.*, 1956, **187**, 637.

Uttal, W. R. Inhibitory interaction of responses to electrical stimuli in the fingers. *J. comp. physiol. Psychol.*, 1960, **53**, 47–51.

Weitz, J. Vibratory sensitivity as affected by local anesthesia. *J. exp. Psychol.*, 1939, **25**, 48–64.

Wilcott, R. C. On the role of the epidermis in the production of skin resistance and potential. *J. comp. physiol. Psychol.*, 1959, **52**, 642–649.

(Received June 23, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 196–199

A CONTRAST EFFECT IN DIFFERENTIAL CONDITIONING[1]

GORDON H. BOWER

Stanford University

Contrary to the interests of parsimony in theory construction, an increasing number of experiments have shown that the reinforcing effect of a given amount of reward is not a static parameter; rather, the conception that emerges from these studies is that the reinforcing effect depends upon the context in which that reward quantity occurs; in particular, it depends upon the range of alternative rewards *S* has received in the experimental situation.

The most direct evidence for this context effect comes from studies of visual discrimination learning. Meyer (1951) and Schrier and Harlow (1956), working with monkeys, and Lawson (1957), working with rats, found that rate of error elimination in a series of visual discrimination problems was not significantly related to the amount of reward when *S* received the same amount (large or small) on every problem. By comparison, if *S* received different amounts of reward for performing correctly on different problems of the learning series, learning rate on a given problem varied directly with the amount of reward for that problem.

These prior studies investigated the context effect in selective learning situations in which the performance measure (percentage of correct responses) presumably reflects the resolution among several competing reaction tendencies. The purpose of the present experiment was to study the context effect in an instrumental learning situation where the presumed relationship between the performance measure (i.e., response speed) and net incentive value is relatively simple. The design of this experiment involved one experimental and two control groups. The experimental (Contrast) *S*s were trained concurrently in two distinct straight runways, with a large reward in one runway (S^+) and a small reward in a different runway (S^-). One control group received a large reward on every trial and a second control group received a small reward on every trial. Asymptotic performance of the Contrast group in S^+ was compared with that of the large reward control group; performance in S^- was compared with that of the small reward control group. A contrast effect would be indicated by significant differences in either or both of these comparisons.

METHOD

Subjects.—The *S*s were 30 male albino rats, approximately 120 days old. They were housed in individual living cages. Taming and adjustment to a 12 gm. feeding diet was carried on for two weeks prior to runway training. The 30 *S*s were randomly assigned to three groups of 10.

Apparatus.—The training apparatus was a double runway similar to that described by Logan, Beier, and Ellis (1955). Two enclosed 5-ft. runways were constructed side by side, the left one painted black and the right one painted white. A single gray startbox adjoined the front end of the apparatus; by closing guillotine doors the entrance to either runway could be blocked. One Standard Electric clock timed the interval from the opening of the startdoor to *S*'s breaking a

---

[1] The assistance of Charles Hoye in the collection of the data is gratefully acknowledged. This research was supported in part by Research Grant M-3849 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

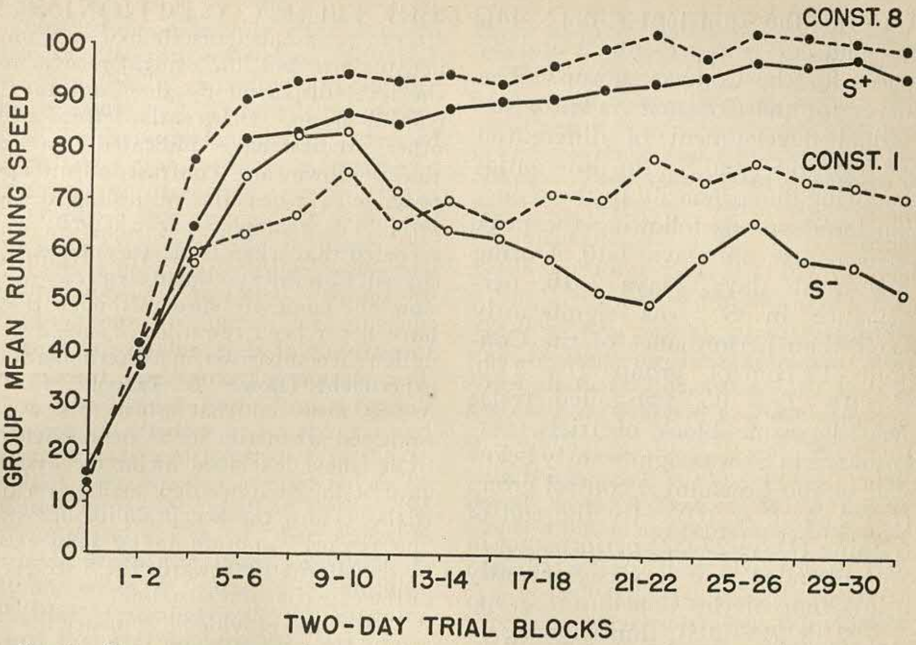


FIG. 1. Group average running speed plotted in blocks of two days. (The dashed curves represent performance of the Constant reward control groups; the solid curves represent performance of the Contrast group in S$^+$ and in S$^-$.)

photobeam 6 in. out in either alley (starttime); a second clock recorded the interval from the first photobeam to a second photobeam located 2 in. from the end of the runways (giving running time). The goalboxes, comprising the last 12 in. of the runways, were equipped with guillotine doors which were lowered after *S* interrupted the goalbox photobeam. Food pellets were placed in pill caps secured to the floor at the end of each goalbox. Between trials, *S*s were detained in a 12 × 16 × 5 in. unpainted wooden box where water was always available.

Procedure.—The Contrast *S*s received half of their four daily trials in one runway (S$^+$) with an eight pellet reward (each pellet weighing 45 mgm.) and the remaining daily trials in the other runway (S$^-$) with a one pellet reward. For half the *S*s, S$^+$ was the left-black alley; for the other *S*s, S$^+$ was the right-white alley. The six possible orders of two S$^+$ and two S$^-$ trials occurred equally often; the daily order for half the *S*s was randomly selected (e.g., + − − +) and the remaining *S*s had the reverse sequence (− + + −). One control group, Constant 8, received four daily trials each rewarded with eight pellets; the Constant 1 control group received four daily trials with a one pellet reward. Half of each control group was run in the right-white runway and half was run in the left-black runway, a given *S* always running in the same runway. The *S*s were left in the goalbox 15 sec., although these times were a bit longer on the first few trials before *S*s ate promptly.

A total of 32 acquisition days (128 trials) were given, half of these to S$^+$ for the Contrast group. Each *S* received its four daily trials in alternation with a second *S*; one rested in the detention cage with water available while the other received one trial; then the first *S* was given another trial, and so on until both had had four trials. Therefore, the intertrial interval varied around 1 min. Ten minutes after the day's trials all *S*s were fed 12 gm. minus the amount of food received in the alley.

RESULTS

The results are presented graphically in Fig. 1 which shows group average running speeds plotted in two-day blocks.[2] The dashed curves represent the Constant groups and the solid curves represent S^+ and S^- speeds for the Contrast group. The curves for the Contrast *S*s show the gradual development of differential behavior to S^+ and S^-, the differentiation being due primarily to the reduction in S^- speeds following the peak performance on Days 7-10. During these peak days, Days 7–10, performance in S^- was significantly faster than performance of the Constant 1 control group ($t = 2.00$, $df = 18$, $P = .06$, two-tailed test); over the same block of trials, performance in S^+ was significantly below that of the Constant 8 control group ($t = 2.84$, $P = .02$). By the end of training (Days 27–32) performance in S^- had fallen to a level significantly below that of the Constant 1 group ($t = 2.48$, $P = .03$), thus showing an ultimate contrast effect in S^-. At the termination of training, performance in S^+ had increased to about the level of the Constant 8 group ($t = 1.19$ for the difference over Days 27–32); both of these speed averages differed from that of the Constant 1 *S*s at well beyond the .01 level.

---

[2] Although starting speeds showed the same pattern of results as did running speeds (somewhat amplified, in fact), they are not pre-

DISCUSSION

These results have shown that the net incentive produced by a small reward is diminished when that reward occurs in a situation where *S* sometimes receives larger rewards. One question raised but not answered in this study is whether this reduced incentive value is a relatively permanent effect; conceivably, after more extended discrimination training, performance in S^- might recover to the level displayed by the Constant 1 control group. Observations from two other studies have indicated that a similar downward contrast effect appears even when the reward in S^- is zero. For example, Logan (1960) has reported that when trials are sufficiently spaced, rats will run down a runway at a slow but moderate speed although they have never been rewarded in such a situation; in comparison, in a discrimination experiment (Bower & Trapold, 1959) it was observed that when rats were rewarded when run in S^+ on half their trials, they developed an active avoidance of the nonrewarded goalbox on S^- trials. Thus, the acceptability of even "no reward" appears to be influenced by the alternative rewards the *S* receives in the experimental situation.

These data provide no evidence for an "upward" contrast effect (i.e., an S^+ performance superior to that of the large reward group); in fact, the data show S^+ performance to be slightly below that of the Constant 8 group. However, it might be argued that the use of a very large reward (eight pellets) precluded the appearance of the upward contrast effect because of the ceiling inherent in the speed measure; conceivably, the effect might appear at lower reward values and this possibility remains to be tested.

One may note that the pattern of contrast effects found here agrees with the findings of studies involving shifts in reward quantity. Reduction in reward quantity frequently leads to a momentary depression in performance of the shifted group to below that of a constant small reward group; in comparison, if sufficient preshift training has been given, increases in reward quantity do not produce a temporary "elation" of the shifted group to above the performance of a constant large reward group.

It is suggested that "downward" contrast effects like those reported here may be understood in terms of a frustration construct such as Amsel (1958) has proposed. Amsel treated only the case

---

sented since they are probably biased in favor of faster starting speeds for the control groups. The control *S*s always ran in the same runway (e.g., left) and undoubtedly learned favorable orienting behaviors in the startbox; the Contrast *S*s, running equally often to the right and left runways, could not learn such favorable preparatory behavior. This differential preparatory factor is quite unlikely to have affected the running times which were recorded after the response was well under way.

where frustration is elicited by nonreward after sufficient prior rewards have resulted in the development within the situation of a stable anticipatory goal response (r_g). To handle the present data, his hypothesis must be extended to include within the class of frustrating events the occurrence of a reward that is smaller than that appropriate to the ongoing r_g amplitude. More direct evidence bearing on such an extension has been reported elsewhere (Bower & Stocks, 1960).

Applying this analysis to the present experiment, the lowered performance in S^- is considered to result from a conflict between anticipation of reward (r_g) and the anticipation of frustration in the S^- goalbox. Frustration is expected to be elicited in the S^- goalbox during the intermediate stages of discrimination training when the large amplitude r_g established in S^+ occurs in S^- through stimulus generalization. It may be noted that this analysis does not imply an upward contrast effect in S^+ and is thus consistent with that part of the data.

SUMMARY

This study investigated the effect on runway performance of rats of giving contrasting amounts of reward in two concurrent learning situations. Rats in the Contrast group received a large reward when run in one alley (S^+) and a small reward when run in a different alley (S^-), the two types of trials occurring in random alternation. Performance in S^+ and S^- was compared with that of two control groups, one receiving a large reward and one receiving a small reward on every trial. Asymptotic performance of the Contrast group in S^- was inferior to that of the small reward control group, whereas performance in S^+ was slightly but not significantly lower than that of the large reward control group. Thus, the effectiveness of a small reward is diminished when it occurrs in a context where *S* sometimes receives larger rewards. These results were interpreted in terms of frustration elicited by the small reward when *S* expects a large reward.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BOWER, G. H., & STOCKS, D. Graded frustration effect as a function of graded reductions in reward. Paper read at meeting of the Psychonomic Society, Chicago, 1960.

BOWER, G. H., & TRAPOLD, M. A. Magnitude of reward and performance in a single presentation discrimination. *J. comp. physiol. Psychol.*, 1959, **52**, 727–729.

LAWSON, R. Brightness discrimination performance and secondary reward strength as a function of primary reward amount. *J. comp. physiol. Psychol.*, 1957, **50**, 35–39.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

LOGAN, F. A., BEIER, E. M., & ELLIS, R. A. The effect of varied reinforcement on speed of locomotion. *J. exp. Psychol.*, 1955, **49**, 260–266.

MEYER, D. R. The effects of differential reward on discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 268–274.

SCHRIER, A. M., & HARLOW, H. F. Effect of amount of incentive on discrimination learning by monkeys. *J. comp. physiol. Psychol.*, 1956, **49**, 117–122.

(Received June 24, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 200–202

T–MAZE REVERSAL FOLLOWING DIFFERENTIAL ENDBOX PLACEMENT[1]

JAMES R. ISON[2] AND DAVID BIRCH

University of Michigan

Discrimination reversal studies by Reid (1953), Pubols (1956), and Capaldi and Stevenson (1957) using choice apparatus and by Birch, Ison, and Sperling (1960) using the method of single stimulus presentation have demonstrated greater ease of reversal for a group given extended training on the original problem than for a group reversed at criterion. Interpretations of this finding have emphasized either the learning of facilitating choice point discrimination responses (Pubols, 1956; Reid, 1953) or the differential extinction of the response to the formerly positive stimulus (Birch, et al., 1960; Capaldi & Stevenson, 1957).

The present experiment is a further investigation of the overlearning reversal phenomenon in which the learning of possibly facilitating choice-point responses during initial training was eliminated. To accomplish this both criterion and overlearning groups were given the appropriate rewarded and nonrewarded experiences by direct placement into detached black and white endboxes. In this initial training there was no exposure to the choice point of the T maze. The endboxes were subsequently attached to the arms of a single unit, mid-grey, T maze and the two groups were given T-maze training. In this training the reward contingency of the endboxes was reversed, the endbox in which *S* received food in placements was now empty and the endbox which was empty in placements was baited.

METHOD

Subjects.—The *S*s were 14 male and 14 female hooded rats, approximately 100 days old at the beginning of training. The male rats were obtained from Rockland Farms, New City, New York, and the female rats were obtained from the laboratory colony maintained by the Psychology Department of the University of Michigan. Three females were discarded, one in each group for failure to run in the T maze and one in the extended training group for failure to eat during the placements.

Apparatus.—A single-unit T maze was used. The stem and both arms were 18 in. long, 4 in. wide, 6 in. high and painted mid-grey. The first 6 in. of the stem served as a startbox and was separated from the rest of the stem by a guillotine type retrace door. Retrace doors were also present at the entrance to each arm and to the endboxes. Baffles placed in the arms prevented reception of the endbox stimuli from the choice point. The endboxes measured 9 in. long, 6 in. wide, and 6 in. high. Four were available, two painted black and two white. The maze was covered by hardware cloth painted the appropriate color. It was maintained in a stable position in a heterogeneous room offering many extramaze cues.

Procedure.—The *S*s were placed on a 23-hr. deprivation schedule by feeding for 1 hr. each day for 14 days prior to the experimental treatments. On each of these days *S*s were handled a minimum of 5 min. They were then assigned to either a 200-placement group (Group 200) or a 50-placement group (Group 50). During initial training *S*s received 10 placements each day, 5 rewarded in the white endbox and 5 nonrewarded in the black endbox with the sequence determined by Gellerman orders (1933). A small metal foodcup was attached to the side of each endbox. The *S* was placed on the side opposite the foodcup with its back towards it. The reward was a 0.05-gm. pellet of dog chow. During the placements the endboxes were removed

---

[1] This research was supported by a grant from the National Science Foundation (NSF-G9779) to the second author.

[2] Now a Rackham Postdoctoral Fellow at the State University of Iowa.

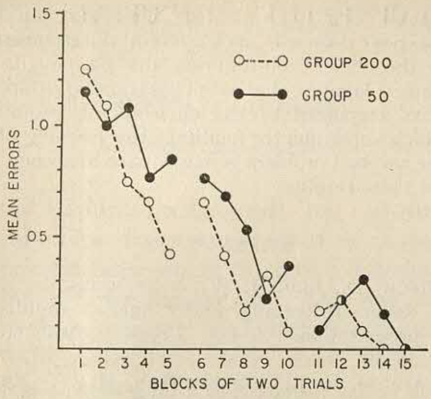


FIG. 1. The mean number of errors in blocks of two trials over the first three days of T-maze training.

from the T maze and their position in the room bore no consistent relation to their subsequent position during the T-maze training. The intertrial interval was 5 min.

The T-maze training was given in 10 noncorrection trials each day to a criterion of 19 correct in two successive days. On the first trial 2 unbaited white endboxes were in place, one on each arm. On all subsequent trials a baited black endbox was placed on the side opposite the first choice. The intertrial interval was approximately 15 min.

RESULTS

Group 200 reached criterion with a mean of 5.92 errors (*SD* = 2.18) and Group 50 with a mean of 8.00 errors (*SD* = 1.88). This difference is significant at the .05 level ($t = 2.44$, $df = 23$). Group 200 reached criterion in a mean of 1.92 days (*SD* = 0.95), Group 50 in a mean of 2.54 days, (*SD* = 0.93). This difference is not significant, ($t = 1.58$, $df = 23$, $P < .20$).

Complete data on the errors in each precriterion trial are available only over the first 3 days of T-maze training. (Several *S*s reached criterion at this point.) The mean numbers of errors in sets of two trials are presented in Fig. 1. Group 200 made fewer total errors over the 3 days. In addition, the within-day distribution of errors appeared to be different for the two groups with Group 200 making relatively more errors in the initial trials and relatively fewer errors in the terminal trials of each day.

For statistical analysis over the 3 days, the 10 trials of each day were grouped into the first 5 trials and the last 5 trials. This analysis is presented in Table 1. In the first 5 trials Group 50 had a mean of 4.46 errors, and Group 200 a mean of 4.50 errors. In the last 5 trials Group 50 had a mean of 3.46 errors, and Group 200 a mean of 1.83 errors. Both Groups and Trials were significant main effects, and the G × T interaction was also significant.

DISCUSSION

One explanation of the overlearning-reversal effect is that, during the extended training, responses are learned to the choice point which facilitate the subsequent discrimination; these responses might be, e.g., the receptor orientations described by Wyckoff (1952). In the present experiment neither group was presented with the choice point stimuli prior to the test problem, but the group given the extended endbox training learned in fewer trials. The differential transfer of facilitating choice point responses from the original problem into the reversal problem is not an

TABLE 1

ANALYSIS OF VARIANCE OF ERRORS MADE IN TRIALS 1–5 AND TRIALS 6–10 OVER THE FIRST 3 DAYS OF T-MAZE TRAINING

Source	*df*	*MS*	*F*
Between *S*s	(24)		
Groups (G)	1	7.88	6.09*
Error (b)	23	1.30	
Within *S*s	(25)		
Trials (T)	1	40.50	29.52**
G × T	1	8.67	6.32*
Error (w)	23	1.37	

* $P = .05$.
** $P = .01$.

adequate explanation of the present results.

A second interpretation of the phenomenon is that the two groups differ in the resistance to extinction of the response to the formerly reinforced stimulus. The design of the present experiment did not permit a direct assessment of the change in strength of the approach response to the previously reinforced goalbox. However, the results are consistent with such an interpretation.

SUMMARY

This experiment tested one explanation of the finding that a group given extended training on a discrimination problem learns its reverse in fewer trials than a group reversed at criterion. Two groups of rats received rewarded placements into a white endbox and nonrewarded placements into a black endbox, one group receiving 50 placements and the other group 200. The endboxes were then attached to the arms of a grey T maze and *S*s received T-maze training. In the T maze the white endbox was empty and the black endbox contained food. The group given the 200 placements learned the T maze with fewer errors. Neither group had any experience with the choice-point stimuli prior to the test problem and could not have learned any differential responses to the choice point during the placements. The explanation that extended training groups learn responses to the choice-point stimuli which subsequently facilitate the learning of the reversal problem is inadequate to account for these results.

REFERENCES

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

GELLERMAN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 207–208.

PUBOLS, B. H. The facilitation of visual and spatial discrimination by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

WYCKOFF, L. B. The role of the observing response in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received June 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 203–205

EFFECTS OF VARIABLE MAGNITUDE OF REWARD ON A LEVER–PULLING RESPONSE[1]

DONALD J. LEWIS AND CARL P. DUNCAN
Louisiana State University *Northwestern University*

"Variable magnitude of reward" refers to the operation of changing reward between at least two different magnitudes for each *S*. One reward is given on each trial, but the magnitude varies irregularly from trial to trial. Using this procedure with rats, Logan, Beier, and Kincaid (1956) found that variable magnitude of reward increased resistance to extinction. Wike, Kintsch, and Gutekunst (1959) have since confirmed this result.

The purpose of the present experiment was to study the effect of variable magnitude of reward on human *S*s, using a lever-pulling response.

METHOD

Apparatus.—The apparatus used in this experiment has been described in detail in previous reports (Lewis & Duncan, 1957, 1958). Basically, it is a modified electrical slot machine. The *S* puts a chromium plated disc into a slot, pushes some buttons, and pulls a lever. If the trial is a rewarded one, *S* receives another disc through a payoff slot; otherwise no payoff is given. The *S* is given a large number of discs and is told that he can play as long as he wishes. He is also told that, at the end of play, he will receive money for each disc that he has won through the payoff trough. He will receive either 1 or 50 cents depending on the condition to which *S* is assigned, and he is informed of the exact amount on each winning trial.

Design and procedure.—All *S*s were given nine acquisition trials; these were the trials over which rewards were distributed. Following acquisition, the extinction trials began and no further rewards were ever given. Eight different groups were run in this experiment. Table 1 summarizes the experimental conditions. The table also shows which trials were rewarded for the various groups and on what trial the reward was given. It can be seen that Group 8 was a 0% reward group; with no money given on any trial, the distinction between acquisition and extinction is purely arbitrary.

Instructions.—The instructions for this experiment were very much like those used for other experiments in this series. The *S*s were told how to operate the machine and at the conclusion of each winning play they were told what the disc was worth, either 1¢ or 50¢. On a nonwinning trial no disc at all was presented to *S*. They were to keep the discs they won separate from those they were playing with and to turn in the former at the close of the experiment for the appropriate amount of money. The *S*s were also told that they could play as long as they liked and that when they decided to quit they should exchange the discs they had won for money. After *S* had decided to quit, he was paid off with coins, given a plausible, but false, explanation of the machine, and told not to tell any other person how the machine operated or how much he had won.

TABLE 1

DESIGN OF THE EXPERIMENT

Trials	Groups							
	1	2	3	4	5	6	7	8
1	1¢	50¢	1¢	50¢	0¢	0¢	1¢	0¢
2	1	50	1	50	0	0	1	0
3	1	1	1	50	0	0	1	0
4	1	50	50	50	1	50	1	0
5	1	50	1	50	0	0	1	0
6	1	50	1	50	0	0	1	0
7	1	1	1	50	0	0	1	0
8	1	1	1	50	0	0	1	0
9	50	50	50	50	1	50	1	0

Subjects.—The *S*s were 480 students from introductory psychology classes at Northwestern University. Sixty *S*s were assigned in sequence as they appeared in the laboratory to each of the eight groups.

---

[1] This study was supported by a grant from the National Science Foundation.

Results and Discussion

The total number of plays to quitting (extinction) composed the basic data of this study. A common logarithm transformation was made for each of these scores and all analyses were performed on these transformed scores.

Three different *patterns* of reward variability were used in this experiment. These were given to Groups 1, 2, and 3 as described in Table 1. To determine whether variability of reward magnitude resulted in increased resistance to extinction the data of Groups 1, 2, and 3 were combined (all had 100% reinforcement), and the mean of these three was compared by t test with the combined means of Groups 4 and 7, the constant 100% groups. Each of the variable reward 100% groups was also compared separately with each of the constant 100% rewarded groups. None of the t tests approached significance.

An analysis of variance was performed on the three variable 100% reward groups taken together. The resulting F was 2.96 with 1 and 77 *df*. This falls considerably short of the value required at the 5% level of significance.

The conclusion to be drawn from these analyses is that variable reward does not increase resistance to extinction in this situation. Further, the three patterns of reward variability manipulated here make no statistically significant difference. It had been expected that Group 1, for which the first eight acquisition trials were rewarded by 1¢ and the final acquisition trial rewarded by 50¢ would be least resistant to extinction. The final, large reward should have made the onset of extinction easier to discriminate. Table 2 shows this apparently to be the case, although the level of significance is inadequate. It was also predicted that Group 3 would exhibit greater resistance to extinction than Group 2. This was because there was a greater number of 50¢ winning trials in Group 2 and the 50¢ trials would stand out more clearly from extinction. Again the predicted trend did occur but the F of 2.96 ($P > .05$) indicates that the interpretation of these differences must be extremely guarded.

TABLE 2

MEAN ANTILOGS OF LOG PLAYS TO EXTINCTION AND OF LOG *SD* IN THE EIGHT GROUPS

Group	Antilog of Mean Log	Antilog of Log *SD*
1	21.23	3.48
2	23.77	3.18
3	35.98	3.15
4	23.66	5.08
5	47.31[a]	2.67
6	47.31[a]	2.38
7	32.14	2.94
8	52.97	2.01

[a] Identical values.

Another question that can be asked of these data has to do with the effect of variability of reinforcement with percentage of reinforcement below 100% held constant. Group 5, for example, has reward variability ranging from 0¢ to 1¢ within a 22% schedule, and Group 6 has reward variability ranging from 0¢ to 50¢, also within a 22% schedule. The means of these two groups did not differ. Within 100% reinforcement, Groups 4 and 7 also afford a test of the effect of constant magnitude of reward on extinction in that for Group 4 the reward is 50¢ on all trials and for Group 7 the reward is 1¢ on all trials. The t tests for the comparison of these two groups yielded a value of 1.20, also short of significance.

All comparisons of the partially reinforced groups with the continuously reinforced groups showed that the partially reinforced groups were more resistant to extinction, supporting evidence from the previous studies in this series (Lewis & Duncan, 1956, 1957, 1958). That the effect of partial reinforcement is clearly present lends indirect weight to the conclusion that varying magnitude is not an effective variable in this situation.

In conclusion it may be remarked that, although this study provides no evidence that variability of reward increases resistance to extinction, there is consistent but nonsignificant evidence that the pattern of reward variability may affect resistance to extinction in that the pattern giving the greatest stimulus change from acquisition to extinction resulted in the fewest plays to extinction. It is surprising that the manipulated variables had so little effect. The *E*s were somewhat sceptical of the lack of significance and continued to run *S*s until they were convinced.

SUMMARY

Fifty *S*s were assigned to each of eight groups, and given a nine-trial acquisition series on a modified "slot machine." Some groups received 100% reinforcement with 50¢ *or* 1¢ on each trial. Other groups received 50¢ and 1¢ irregularly on the acquisition trials—conditions of "variable magnitude of reward." Other groups received partial reinforcement. At the end of acquisition, rewards ceased, but *S*s were allowed to play as long as they liked. The basic response measure was the number of responses to quitting (extinction). Partial reinforcement had its usual effect in retarding extinction, but there was no evidence that variability of reward affected extinction in any way.

REFERENCES

LEWIS, D. J., & DUNCAN, C. P. Effect of different percentages of money reward on extinction of a lever-pulling response. *J. exp. Psychol.*, 1956, **52**, 23–27.

LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and amount of reward. *J. exp. Psychol.*, 1957, **54**, 115–120.

LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55**, 121–128.

LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

WIKE, E. L., KINTSCH, W., & GUTEKUNST, R. The effect of variable drive, reward, and response upon instrumental performance. *J. comp. physiol. Psychol.*, 1959, **52**-403–407.

(Received June 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 206–207

SUPPLEMENTARY REPORT: INTERPOLATED UCS TRIALS IN GSR CONDITIONING

R. A. CHAMPION

University of Sydney

Dufort and Kimble (1958) have reported that UCS-alone trials led to the same performance increment as CS-UCS trials when no ready signal was given before each trial, but others (Goodrich, Ross, & Wagner, 1959; McAllister & McAllister, 1960) were unable to obtain this effect and found that interpolated UCS trials rather led to a decrement in performance. The following experiment was designed to determine if the effect would occur in the conditioning of the GSR. Although the nature of the GSR precludes the use of a ready signal, which would itself elicit a response, the experiment nevertheless allowed a clear test of the conflicting findings.

Method

Subjects.—The *S*s were 30 volunteers from courses in psychology at the University of Sydney. They were allocated alternately to the two groups of the experiment in the order of their appearance at the laboratory. Two *S*s were discarded, one from each group, because they failed to give any responses on test trials; this left 7 men and 7 women in Group S and 9 men and 5 women in Group TS.

Apparatus.—The apparatus[1] was identical with that described elsewhere (Champion & Jones, 1961) except that (*a*) the CS was a 2000-cps, 75-db., sine-wave tone of 2000-msec. duration presented to *S* through headphones, and (*b*) the onset of the CS occurred 1000 msec. before the onset of the UCS (shock).

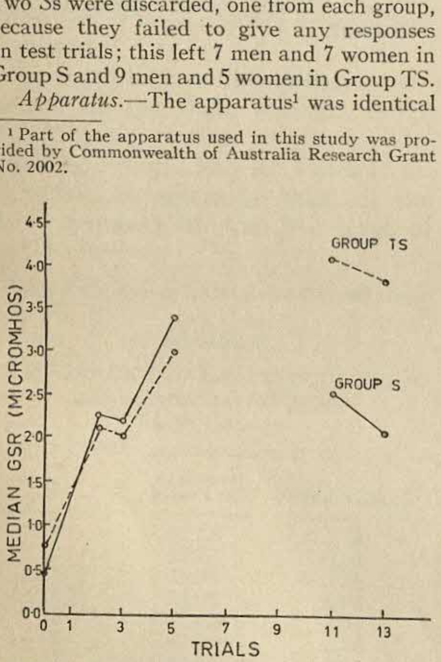


FIG. 1. Median GSR as a function of the number of trials before each test trial in groups in which Trials 6–10 were UCS-alone (Group S) and CS-UCS (Group TS).

Procedure.—The first 10 trials were the same for all *S*s and consisted of the following sequence of CS-alone trials (T) as tests of performance, UCS-alone trial (S), and CS-UCS trials (TS): T, S, TS, TS, T, TS, T, TS, TS, T. For Group S the next 5 trials were UCS-alone trials and for Group TS they were CS-UCS trials. The last 5 trials in the procedure were again the same for both groups and made up the sequence TS, T, TS, TS, T. Thus tests of performance through CS-alone trials were made after 0, 2, 3, 5, 11, and 13 CS-UCS trials or UCS-alone trials without taking into account Trial 2, shock alone. After the first two trials no instructions were given when the stimulus conditions were changed.

Results

The obtained performance curves are shown in Fig. 1. Because the distribution of scores was not normal the median GSR was used as a dependent variable and nonparametric statistical tests were applied to the data. The interpolated UCS-alone trials produced a decrement in performance in Group S whereas the corresponding CS-UCS trials produced an increment in Group TS. A U test applied to the scores on Trials 11–13 (Fig. 1) showed the difference between Groups S and TS to be significant ($z = 1.95$, $P = .05$). The decrement in Group S and the increment in Group TS were examined separately by applying rank tests to a comparison of the total score on the two test trials before and after the interpolated UCS-alone trials or the corresponding CS-UCS trials, i.e., Trials 3–5 vs. Trials 11–13 in Fig. 1. The respective z values were 1.70 ($P = .09$) and 1.67 ($P = .05$ with a one-tailed test).

It is clear that the interpolated UCS trials in Group S did not have the same effect as the normal training trials in Group TS under the conditions of this experiment and with

---

[1] Part of the apparatus used in this study was provided by Commonwealth of Australia Research Grant No. 2002.

no ready signals. This finding casts doubt on the conclusion of Dufort and Kimble (1958) that the absence of a ready signal is the critical factor in allowing UCS-alone trials to produce the same effect as conditioning. On the other hand the similarity of the present finding to those of Goodrich et al. (1959) and McAllister and McAllister (1960) is striking when the well known differences between the eyelid response and the GSR are taken into account.

REFERENCES

CHAMPION, R. A., & JONES, J. E. Forward, backward, and pseudoconditioning of the GSR. *J. exp. Psychol.*, 1961, **62**, 58–61.

DUFORT, R. H., & KIMBLE, G. A. Ready signals and the effect of interpolated UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **56**, 1–7.

GOODRICH, K. P., ROSS, L. E., & WAGNER, A. R. Supplementary report: Effect of interpolated UCS trials in eyelid conditioning without a ready signal. *J. exp. Psychol.*, 1959, **58**, 319–320.

MCALLISTER, W. R., & MCALLISTER, D. E. The influence of the ready signal and unpaired UCS presentations on eyelid conditioning. *J. exp. Psychol.*, 1960, **60**, 30–35.

(Received April 27, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 2, 207–208

SUPPLEMENTARY REPORT: FAMILIARITY OR LOW VALIDITY?

E. JAMES ARCHER[1]

University of Wisconsin

In a recent article Lindley (1960) concluded, "that familiarity is an important variable in serial nonsense syllable learning. Even though the published association values of the syllables . . . were equal, the familiar syllables were significantly easier to learn than the unfamiliar syllables" (p. 369). It is possible that familiarity and association value are independent characteristics of verbal materials, but it is also possible that Lindley's results and conclusions are a function of the invalid and unreliable association values described by Glaze (1928) and Krueger (1934). Even if one ignores the very probable change in association value over a 30-yr. period, one cannot ignore the fact that there were only 15 *S*s in the Glaze study and each of Krueger's *S*s responded to only about 200 items, i.e., individual differences among *S*s were ignored.

On the assumption that the recent tables of association value published by Archer (1960) are more reliable and valid (each of 216 *S*s responded to all 2480 items) the lists which Lindley used were re-evaluated. The lists are reproduced in Table 1 and show the Glaze and Krueger values, where appropriate, as well as the Archer values.

As may be seen there are considerable discrepancies. Of immediate interest are the pairs of mean values for Lists 2G and 3G and Lists 2K and 3K. These lists were thought by Lindley to have equal association

TABLE 1

LISTS OF TRIGRAMS USED BY LINDLEY AND THEIR ASSOCIATION VALUES(%)

List 2G			List 3G			List 2K			List 3K		
Syl.	A[a]	G[b]	Syl.	A	G	Syl.	A	K[c]	Syl.	A	K
WID	65	80	RUJ	18	93	WES	84	91	LUV	88	96
FOL	65	93	WIM	85	87	BUR	91	92	QIT	66	96
JAC	89	87	BOC	59	87	PIC	99	96	ROX	91	95
KEP	70	87	DEY	68	87	GOV	95	95	WUN	81	96
BIR	81	87	LAQ	58	73	NAM	73	85	DAC	62	89
HOM	91	100	VIF	21	100	JUL	81	93	YEZ	54	95
SUG	47	87	TAZ	32	67	DIF	79	97	KOS	54	95
NEX	61	73	NOK	75	—[d]	YEL	92	99	JEM	90	94
GAV	77	67	HUP	75	80	TAK	88	99	FAX	59	91
Mean	72	85		55	84		87	94		72	94

[a] Association values from Archer (1960).
[b] Association values from Glaze (1928).
[c] Association values from Krueger (1934).
[d] Glaze (1928) reports two values for NOK; 87% and 53%.

---

[1] Paper prepared while author was an NSF Senior Postdoctoral Fellow at the Applied Psychology Research Unit, Cambridge, England.

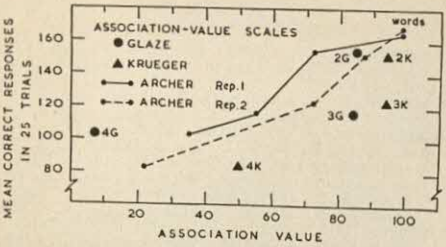


FIG. 1. Lindley's data replotted as a function of Archer's association-value scale. (Each point is the mean of 12 different *S*s.)

values. Where he thought there was only a 1% difference between 2G and 3G, the difference on the Archer scale is 17%; where he thought Lists 2K and 3K to be equal, there is a 15% difference on the Archer scale. The previously assumed association values for Lists 4G and 4K were 7% and 49%, respectively; whereas, the corresponding Archer values are 35% and 22%.

Figure 1 shows Lindley's (1960) data replotted as a function of the Archer association-value scale. Each point is the mean of a group of 12 *S*s (total $N = 96$). In Replication 1, one group learned the list of three-letter words, and each of the other three groups learned one of the "Glaze" lists (2G, 3G, or 4G). In Replication 2, one group learned the list of three-letter words and each of the other three groups learned one of the "Krueger" lists (2K, 3K, or 4K).

The increase in performance (mean correct responses in 25 trials) as a function of association value is much more regular than the Lindley paper indicated. There is no reason to expect the function to be linear, but it is reasonable to expect it to be monotonic. Such seems to be the case. By contrast, the pairs of points 2K–3K and 2G–3G would suggest a double-valued function.

Lindley may still be right about the effect of familiarity but his data do not support the hypothesis.

REFERENCES

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.

LINDLEY, R. H. Association value and familiarity in serial verbal learning. *J. exp. Psychol.*, 1960, **59**, 366–370.

(Received February 3, 1961)

Journal of
Experimental Psychology

VOL. 62, NO. 3 SEPTEMBER 1961

LEARNING WITH REDUCED FEEDBACK INFORMATION[1]

MAYNARD W. SHELLY

Office of Naval Research

The information *S* receives concerning his responses undoubtedly affects the rate at which he learns the correct response. Although this is obvious, two problems remain largely unexplored. These are the relationship between amount of feedback information and initial rate of learning, and the relationship between amount of feedback information and the general shape of the empirical learning curve.

Most investigations on the effects of knowledge of results have simply varied the feedback information through eliminating or reducing the number of events following *S*'s response. This is true of the old line drawing experiments (e.g., Trowbridge & Cason, 1932) and the more recent studies involving contingent and noncontingent feedback (e.g., Bourne & Pendleton, 1958; Neimark, 1956; Shelly, 1960). Feedback information giving the correctness of a response may also be decreased by increasing the number of events following a response, which may be considered to be noise added to the information about the correct response. It is the purpose of the present study to investigate some of the effects of adding such noise to the information fed back to *S* within a paired-associate paradigm.

This paper consists of three experiments, each of which varied some characteristic of the noise added to the information given *S* about the correct response. Experiment I varied the amount of noise which was added to the correct information. In Exp. II the sequential characteristics of the noise over trials was varied, while in Exp. III the distribution of noisy events over the set of available incorrect response alternatives was varied.

Each experiment asked two questions. The first question was whether the reduction in initial learning rate was proportional to the amount by which the noise decreased the information fed back to *S*. (Here, and in the following, a decrease in information refers to a decrease in the knowledge needed to identify the correct response, rather than to a decrease in uncertainty.) The second question was whether the reduction in information produced the same temporal pattern of effects for all conditions. This latter question is part of the question as to whether difficulty of the task alters the shape of the learning curve.

---

[1] The data for these experiments were collected while the author was at the University of Wisconsin. Some of the results together with some results not included in this paper were presented at the Midwestern Psychological Association meeting in St. Louis in 1956. The author wishes to thank Northwestern University for providing him with facilities to complete the analyses of these experiments, the Office of Naval Research for encouraging the professional interests of its scientists, and Carolyn Shelly for helping in the editing of the final draft of the paper.

GENERAL METHOD

Task.—The task employed in these experiments was a slight modification of the usual paired-associate paradigm. The *S* was to learn to associate responses (pulling switches) with visual stimuli. All conditions employed eight stimuli. The stimuli were either angles or modifications of a triangle. In some conditions there were four response alternatives and in others, eight alternatives were available. When there were eight response alternatives each stimulus had a unique response associated with it; when there were only four response alternatives, each response alternative was correct for exactly two stimuli. The stimuli were presented to *S* on a screen placed before him. Whenever a stimulus appeared on the screen *S* was to attempt to select that response alternative which would be correct for the stimulus being presented.

Under conditions with no noise, a lamp would light following the response made by *S*; this lamp, located above some switch, indicated which switch was correct for the preceding stimulus. Under the conditions with noise, more than one lamp would light following a response by *S*. The additional lamp (or lamps) represent what is referred to as noise in the experiments in this paper. Only one of the lamps, however, would *always* come on for a given stimulus; this lamp, the constant lamp, would indicate the correct response for the given stimulus. The other lamps, occasionally lit after a response, will be called noise lamps. Thus, in the conditions with no noise, *S* was required only to associate the correct response with each stimulus. In the conditions with noise, *S* had both to identify the correct response for a given stimulus and associate this response with the stimulus, although these two tasks are not really distinct.

Apparatus.—The apparatus consisted of two parts; an experimental booth in which *S* sat, and the programing and recording equipment exterior to the booth. The interior of the booth was flat black. On a sloping table in the booth, built into an aluminum chassis, were the response switches and the green-domed feedback lamps mounted above their corresponding switches. Those switches not in use were covered by a box. The stimuli were presented by strip film projector, and the feedback programed on Western Union Tape Transmitters.

Subjects.—The *S*s were volunteers from elementary psychology courses at the University of Wisconsin. Each group was approximately equally divided between men and women.

General procedure.—When *S* reported to the laboratory he was, within the restriction that the same number of *S*s serve in each condition, randomly assigned to one of the conditions of the experiment.

In the instructions *S* was told that this was a learning experiment. He was told that one switch would be correct for each different stimulus. He was to learn to pull the correct switch each time another stimulus appeared. If he were assigned to a condition free of noise, he was told that the lamp which came on after he made a response would indicate the correct response for the preceding stimulus. If *S* were in a noisy condition he was told that several lamps would come on after his response, but that only *one* of these lamps would *always* light at the end of a trial; the switch under this lamp was the correct switch. His task, he was told, was to learn which lamp would follow each stimulus just as in everyday life we learn to anticipate consequences which *always* follow a given action even though on any one occasion, many things might happen. A brief demonstration was then given.

The stimuli were presented to *S* at a 4-sec. rate. About .25 sec. before the stimulus changed one or more of the green-domed lamps came on and stayed on until the stimulus changed.

In Exp. I the stimuli were presented in blocks of 80 trials; blocks of 96 trials were used in Exp. II and III. In each block of trials each stimulus occurred equally often (10 times in Exp. I and 12 times in Exp. II and III); the stimuli were assigned their sequential positions randomly without replacement. Three different blocks of stimuli were employed and each block was presented twice. The six blocks composing each experiment were presented in a random order.

Notation for and general characteristics of the experimental conditions.—To make it possible to describe the conditions of this paper succinctly, the following notation will be used. A parenthesis containing either four or eight numbers will be used to represent each experimental condition. Assume reference is being made to a single fixed stimulus. Then, the first position in the parentheses will be reserved for information concerning the response alternative which is correct for this fixed stimulus; the number in this first position will equal the number of times this given stimulus appears in a block of trials. The lamp associated with this position is the constant lamp. The remaining positions denote the number of times a noise lamp will become lit above another alternative following *S*'s responses.

As an example of this notation, in Exp. I there were three conditions: in the first condition no noise occurred, in the second condition exactly one noise lamp occurred, and in the third there were two noise lamps. Then the sum of all numbers in the parentheses beyond the first number will equal zero if there is no noise, will equal the number of times the stimulus appears in a block of trials if one noise lamp occurs with each stimulus, or will equal twice the number of times the stimulus appears in a block of trials if two noise lamps occurred with each stimulus. Thus if the stimulus appears 10 times in a block of trials, possible admissable conditions would be the following: (10, 0, 0, 0), (10, 0, 0, 0, 0, 0, 0, 0), (10, 2, 4, 4), (10, 4, 8, 8), (10, 2, 4, 4, 0, 0, 0, 0,), (10, 2, 2, 2, 2, 2, 0, 0), and (10, 8, 8, 1, 1, 1, 1, 0). The first two representations would be for conditions free of feedback noise having four and eight available alternatives, respectively. The numbers included in the third parentheses represent a condition in which a false or misleading lamp comes on above one wrong alternative twice and above the other two incorrect alternatives four times in a block of 10 trials. To control against biases arising from the placement of alternatives, especially in Exp. II and III, two different associations of correct stimuli with response switches were used for each condition.

Measure of information.—Since the purpose is to investigate some of the effects of decreasing information in a learning situation, it is necessary to employ some measure of information. Except for Exp. II, where an additional measure of information is used, the measure of information was computed as follows. For each stimulus in the sequence of stimuli presented to *S*, determine the expected (average) number of times this stimulus must be presented to an *S* with perfect memory before he can identify the correct switch for that stimulus. Take the logarithm to the base two of the number of response alternatives available to *S*. Divide this number by the expected number of trials required to identify the correct switch. This value will be designated by an I. I values can be viewed roughly as the average number of bits presented to *S* per trial. When the number of alternatives is fixed, this value is simply the reciprocal of number of trials required to identify the correct alternative for the fixed stimulus.

EXPERIMENT I

The purpose of Exp. I was to study some effects of the amount of noise upon learning through varying the number of noisy events which followed a response.

Method

Conditions.—There were six conditions. In the following list of conditions, the values in the parentheses, in other than the first position, represent expected values over three blocks of trials:

1. (10, 0, 0, 0), $I = 2.00$
2. (10, 3.3, 3.3, 3.3), $I = .86$
3. (10, 6.7, 6.7, 6.7), $I = .47$
4. (10, 0, 0, 0, 0, 0, 0, 0), $I = 3.00$
5. (10, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4), $I = 1.43$
6. (10, 2.9, 2.9, 2.9, 2.9, 2.9, 2.9, 2.9), $I = .97$

Conditions 1 and 4 were conditions free of noise. In Cond. 2 and 5 one noise lamp, occupying a position selected randomly from among the incorrect positions, appeared on each trial. In Cond. 3 and 6 two noise lamps appeared on each trial; the sequence of pairs of positions occupied by the noise lamps was random in these two conditions.

Subjects and stimuli.—A total of 10 *S*s served in each condition. Four of the *S*s were, however, replacements for *S*s who apparently could not understand the instructions. These 4 *S*s, all in noise conditions, showed no signs of learning anything, and at the end of the experiment could not explain what they were supposed to have done in the experiment or gave erroneous explanations.

The stimuli used were eight angles: 0°, 45°, 90°, 135°, 180°, 225°, and 315°. All were drawn in standard position.

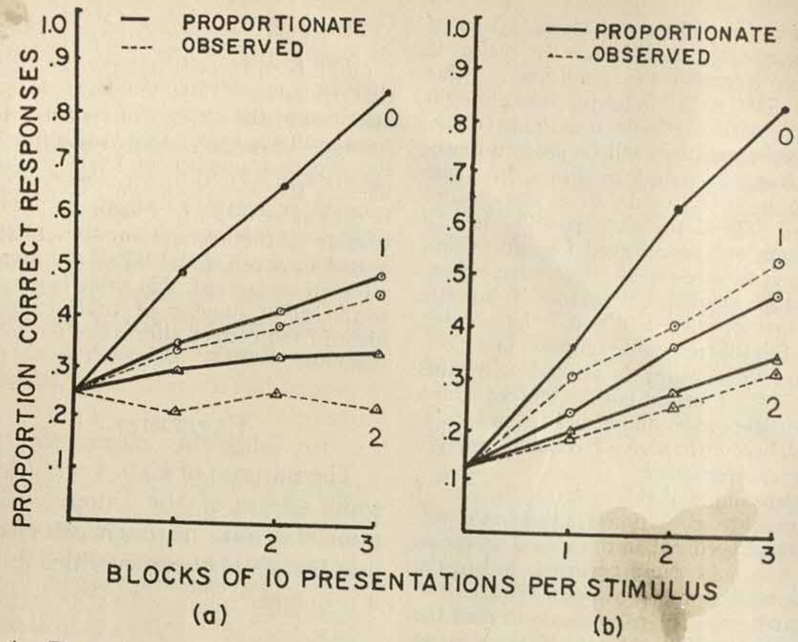


FIG. 1. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials in Exp. I for conditions with four-alternative (a) and eight-alternative (b) responses.

Results

Initial rate of learning.—To examine the effects of reduced feedback information upon initial rate of learning only the data from the first three blocks of trials were used. After the first three blocks, the no-noise groups were approaching perfect performance. This made it difficult to determine the appropriate predicted performance for the noise groups.

In Fig. 1 are presented the data for the first three blocks of trials. The curves in Part a are for four-alternative conditions; those in Part b for the eight-alternative conditions. The curves in Parts a and b labeled "O" are the learning curves for the conditions free of noise. These curves act as both empirical results and the basis for the theoretical predictions of proportionality. Below these curves are both solid and broken curves. The solid line curves below these O curves are computed on the basis of *I* values. They represent the percentage of correct responses which would result had *S*s shown a loss, for a fixed block of trials, proportionate to the decrease in information. That is, they were computed as follows. The value on the O curve was taken for a fixed block of trials. This was then multiplied by the ratio of the *I* value for the curve being computed to the *I* value for the appropriate O curve. The dotted curves represent the obtained results. Data points for each pair of predicted and obtained curves are enclosed by identical symbols. Each set of solid and broken curves are labeled by a number. The numbers represent the number of noise lamps which appear per trial in the conditions upon which the curves are based. As can be observed for both four (Part a) and eight (Part b) alternatives, the correspondence is generally quite fair. For the condition with four alternative responses, the prediction of proportionality agrees with the results within 19.4% error. For the eight-alternative response conditions, the agreement is better, to within 14.0% error.

To obtain a measure of the correspondence between the proportionate predicted values and the obtained values, the following measure was employed. Let X_i be a random variable, normally distributed with mean zero and variance 1. Then define the derived random variable W to be:

$$W = \sum_{i=1}^{N} \tfrac{1}{2}(X_i + X_{i+1}) - c(X_i, X_{i+1}) \frac{X_i X_{i+1}}{X_i + X_{i+1}}$$

where $c(X_i, X_{i+1}) = 0$ if sign $(X_i) =$ sign (X_{i+1}), i.e., both X_i and X_{i+1} have negative values or both variables have positive values; or $c(X_i, X_{i+1}) = 1$ if sign $(X_i) \neq$ sign (X_{i+1}). W then represents the random variable which takes on as values the *area* between the predicted and obtained curves, when the distances between corresponding points on the curves have been transformed into appropriate scores based on standard deviations and successive points along the abscissa are separated by unit distances. The area obtained in this manner will be called a *normalized deviation area.*

W was used as a measure of the amount of correspondence between the predicted and obtained results. This measure was used in preference to others for two reasons. First, this measure takes into account the amount of variability at each data point; small deviations between curves can be quite important if the variability is correspondingly small. Secondly, the use of areas instead of distances between a small number of corresponding points, is sensitive to the sequential ordering of data points about the theoretical points. If the data points lie consistently to one side of the theoretical points, then the normalized deviation areas would be larger than were the two curves to cross several times.

For Experiment I, the normalized deviation areas are 1.97, 8.68, 1.68, and 2.10 for Cond. 2, 3, 5, and 6, respectively.

Shape of the learning curve.—The following analysis is directed towards the effects of reduced information upon the shape of the performance curve. There are many specific questions which could be asked on the general problem of shape of learning curves. One of the most interesting is the location, in the empirical learning curve, of the point at which the greatest increase in the number of correct responses takes place. This of course has a direct implication for the shape of the learning curve. For the usual learning curve, this place of maximum increase would be between the beginning of the experiment and the end of the first block of trials. If it occurred elsewhere, the learning curve would not be monotonically negatively accelerated. The following analysis will be concerned with the maximum increase in the proportion of correct responses between successive blocks of trials; this quantity will be denoted by MaxΔR. The chief interest will be the location of the MaxΔR in the learning sequence. In Fig. 2 the value of MaxΔR is given by the height of the bars. It may be seen that the place at which MaxΔR occurs varies with experimental conditions. The meaning of "Block after which MaxΔR occurs" in Fig. 2 is as follows: If the maximum increase in the proportion of correct responses occurs between (*a*) the proportion of correct responses which could be expected on a chance basis and (*b*) the proportion of correct responses obtained on the first block of trials, MaxΔR is considered to have occurred after block "O." If MaxΔR occurs between the second and third block of trials, it is included in the graph as occurring after Block 2. The last place MaxΔR could occur in this experiment would be between Blocks 5 and 6, or after

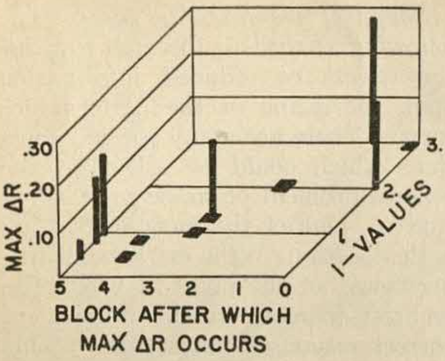


FIG. 2. Amplitude of MaxΔR, and the relation between the block after which it occurs and *I* value of the feedback information for Exp. I.

Block 5 in Fig. 2. This dimension located the *position* of MaxΔR in the learning sequence. The other dimension on the floor of the graph simply gives the *I* values listed previously for the conditions of Exp. I.

Sometimes the differences between the increases in proportions of correct responses between blocks are quite small. It is consequently desirable to establish the reasonableness of two types of statistical hypotheses. First, it is desirable to establish that the distribution of positions of MaxΔR cannot be accounted for by chance deviations from either the zero or first block positions. Secondly, because the relationship of the position of the MaxΔR values to the *I* values is important, it is also desirable to obtain some estimate of this relationship.

The probability of finding a set of deviations in the position of MaxΔR from the zero block position as large or larger than those obtained in this experiment, under the hypothesis of a rectangular distribution, will be denoted by P_0. The probability of finding a similar set of deviations from the Block 1 position as large or larger than those obtained will be denoted by P_1. For this experiment the values of these two probabilities were:

$$P_0 = \frac{15}{6^5} < .001$$

and

$$P_1 = \frac{12}{5^5} < .004$$

It is thus unlikely that MaxΔR occurs during the first part of the learning sequence in all conditions.

Consider a line drawn from the corner of the zero block and *I* value of 3.00 to the diagonally opposite corner, Block 5 and *I* value zero; this line would result from a perfect linear relationship between MaxΔR and *I* value which passed through the two described corners. This line has no theoretical significance. It was chosen because if the relationship between position of MaxΔR and the *I* value approximates this line, it will be clear that the "shape of the learning curve" depends upon the *I* value of the feedback information. The positions that the columns would occupy were there such a perfect relationship in Exp. I between these two variables are indicated by the rectangles on the floor of the graph; these rectangles do not lie solely on the diagonal. The rectangles in Fig. 2 could lie solely on the diagonal for the *I* values used only if MaxΔR could assume values other than whole numbers. Since this is not logically possible within the framework of the experimental analysis, the illustrative rectangles are placed on the *nearest* whole number of the dimension "Block after which MaxΔR occurs." This results in their only approximating a diagonal.

The probability of finding a set of deviations from this hypothetical diagonal line as small or smaller than those obtained, under the hypothesis of a rectangular distribution, will be denoted by $P_{Diag.}$. This gives the probaability that, as graphed, there is not a relationship close to a linear relationship passing through the two corners between the position of MaxΔR and the I values of the conditions of Exp. I. The value of this probability was: $P_{Diag.} = \frac{2}{6^4} < .002$. Therefore the probability that the position of MaxΔR is not closely related to the I values of feedback is small. Consequently, the shape of the learning curve varies with information fed back.

It will be noted that although there seems to be some tendency for the values of MaxΔR to decrease as the I values of the feedback decrease, the size of the decrease is not striking. The tau correlation coefficient between these two variables is, however, .866.

EXPERIMENT II

In Exp. I the I values of the feedback information were changed by varying the number of noisy events which followed a response. Experiment II attempts to extend these results to conditions in which the I values are changed by varying the sequential patterning of the noisy events. Patterning is varied principally by varying the number of times a noise lamp remains in the same position for successive presentations of a given stimulus.

Method

Conditions.—The conditions for Exp. II are presented in Table 1, together with the I values for each condition. For all conditions, except the two noise-free groups, two lamps came on following each of *S*'s responses. One of these lamps, the constant lamp, indicated the correct response; the other was a noise lamp. In all but one of the noise conditions, (Random$_1$), each stimulus was paired with two positions which the noise lamp could occupy. A lamp in one of these two positions would light, together with the constant lamp indicating the correct response alternative, following each presentation of the appropriate stimulus. The effects of the sequential patterning of the position of noise lamps are the particular interest in this experiment.

TABLE 1

CONDITIONS OF EXP. II

Conditions: Frequency or Pattern of Change in Noise Lamp	4-Alternative Conditions		8-Alternative Conditions	
	I_1	I_2	I_1	I_2
No noise (N)[a]	2.00	2.00	3.00	3.00
Alternating (A)	1.00	1.00	1.50	1.50
Random$_1$ (R$_1$)	.78	.78	1.43	1.43
Random$_2$ (R$_2$)	.67	.67	1.00	1.00
By quarter blocks (Q)	.72	.50	1.09	.67
By half blocks (H)	.58	.29	.87	.43

[a] The letters in parentheses denote the symbols by which the conditions are referred to in Fig. 3 and 4.

In the alternating noise conditions (Cond. A), the two noise lamp positions for each stimulus are employed alternatively. After the presentation of a given stimulus a noise lamp in one position lights; on the next presentation of this stimulus the noise lamp in the other position comes on. The position of the noise lamp continues to alternate in this fashion.

In the conditions designated as "By Quarter Blocks" (Cond. Q), one noise lamp position was employed successively for three presentations of a given stimulus; then the noise lamp in the other of the two positions it assumed would come on for the next three presentations of the same stimulus. Thus the noise lamp position which was paired with a given stimulus changed every quarter of a block.

In those conditions labeled "By Half Block" (Cond. H), the noise lamp position changed every half block. The noise light would come on in one position for six presentations of a given stimulus and then the other position of the noise lamp would be employed for the next six presentations remaining in the block. Thus, the noise lamp changed positions at the beginning of a block of trials, and once more half way through the block.

The no-noise conditions are the same as Cond. 1 and 4 in Exp. I, except that each stimulus is presented 12 times in one block.

Two types of random conditions were used in this experiment. Random$_1$, (Cond. [12, 4, 4, 4] and [12, 1.7, 1.7, 1.7, 1.7, 1.7, 1.7, 1.7]), was the same as Cond. 2 and 5 in Exp.

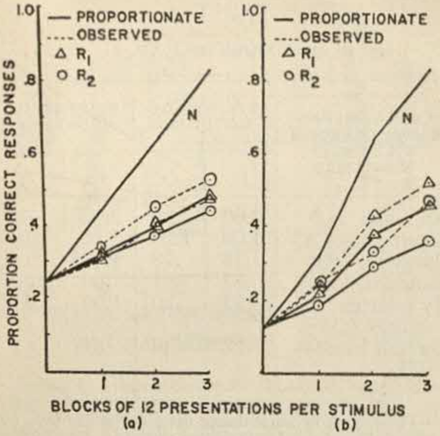


FIG. 3. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials for the two random conditions of Exp. II. (Four-alternative, a, and eight-alternative, b, conditions.)

I with the exception that the number of stimuli per block was changed. Random_2 (Cond. [12, 6, 6, 0] and [12, 6, 6, 0, 0, 0, 0, 0]) acts more closely as a control for conditions of Exp. II in that the noise lamp can assume only one of two incorrect positions. For Random_2 conditions the probability on any one trial of the noise lamp changing from its previous position is .5.

Information measures.—In Table 1, following each condition, two *I* values are given. I_1 is the same information measure as described in the General Method section. It is based on the expected number of trials required for an *S* with perfect memory to identify the correct response alternative, when the beginning point in the sequence is chosen at random. I_2 is identical except that the beginning points in the sequence are considered to occur immediately after a change in the position of a noise lamp. This increases the expected number of trials required to identify the correct response alternative for a fixed stimulus in the Cond. Q and H, and since in Cond. A the position of the noise lamp alternates between the two possible positions, I_1 equals I_2. I_2 is employed in addition to I_1 since it is conceivable that an *S*'s efforts to filter out the noise may begin with or be intensified by the occurrence of a change in the noise lamp's position. There are then two bases by which to compare performance in the noise conditions.

Subjects and stimuli.—Ten *S*s served in each condition. Two of these *S*s were replacements for *S*s who apparently did not understand the instructions. The same criteria for rejection were used as in Exp. I.

The eight stimuli used were clearly discriminable modifications of a solid black isosceles triangle. These stimuli were formed with all possible combinations of the following: the presence or absence of a hole in the center of the triangle; the presence or absence of a cut perpendicular to the base which eliminated one corner; and the presence or absence of a cut oblique to the base which eliminated the other corner at the base of the triangle.

Results

Initial rate of learning.—In Fig. 3 are presented the curves which are predicted on the basis of proportionality and the empirical curves obtained for the first three blocks of trials for the two types of random noise conditions. Part a presents the data for the groups with four response alternatives and Part b for eight alternatives. It can be seen that in general the prediction of a decrease in initial learning rate proportional to the decrease in *I* value is more accurate for the Cond. Random_1 than for Cond. Random_2. The percentage error between predicted and observed is 11.7 for Random_1, while for Random_2 it is 21.3%. That the prediction of proportionality is more adequate for Random_1 is also apparent in Table 2 which presents the normalized deviation areas. Since there was a difference between Random_1 and Random_2 the results to be presented below concerning the sequential patterning of the noise lamp positions should possibly be considered to be limited to conditions in which the noise lamp is restricted to two positions. (The findings of Exp. III will further support the suggestion that the number of incorrect positions a noise lamp can occupy must be considered when evaluating the prediction of proportionality.)

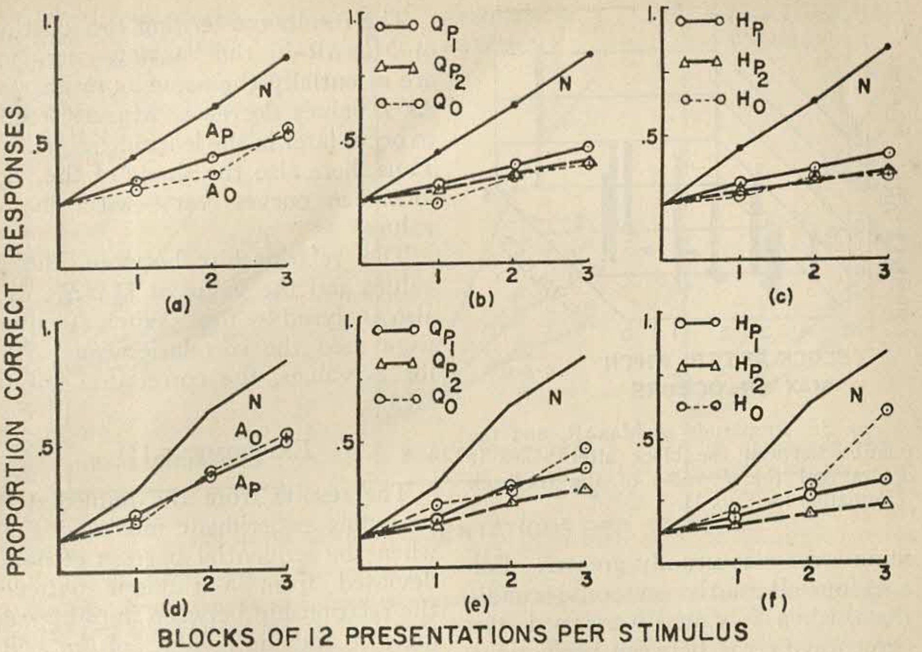


FIG. 4. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials for Cond. A, Q, and H of Exp. II. (For Cond. Q and H, proportionate predictions are made both on the basis of I_1 and I_2. The upper three graphs are for the four-alternative conditions and the lower three for the eight-alternative conditions.)

In Fig. 4 the predicted and obtained curves for Cond. A (Parts a and d), Cond. Q (Parts b and e), and Cond. H (Parts c and f) are presented; the data for the groups with four response alternatives are presented in the upper three graphs. An A, Q, or H with a subscript O refers to the obtained results. The same letters with a *P* subscript refer to the predicted proportionate results; P_1 indicates that the curve was computed using I_1 as a measure of feedback information, while P_2 indicated that I_2 was used. Curves labeled "N" are based on a noise-free condition.

In general it can be observed that the use of I_2 provides better predictions for the conditions with four response alternatives while predictions based upon the I_1 values are superior for the eight-response alternative conditions. This is generally borne out by the normalized deviation areas presented in Table 2, where the normalized deviation areas involving I_2 are included in parentheses.

The percentage error between the proportionality prediction and the obtained results for the alternating conditions was 7.3, for the four- and eight-alternative response conditions combined. For Cond. Q and H, this

TABLE 2

NORMALIZED DEVIATION AREAS FOR EXP. II

Noise Distribution	4-Alternative Conditions	8-Alternative Conditions
Alternating	2.14	.76
Random$_1$	1.51	2.39
Random$_2$	4.42	3.52
Quarterly	4.53 (1.99)[a]	3.14 (7.58)
Halves	3.44 (.85)	4.50 (7.85)

[a] The values in parentheses represent the normalized deviation areas when I_2 values are used as the standard instead of I_1 values.

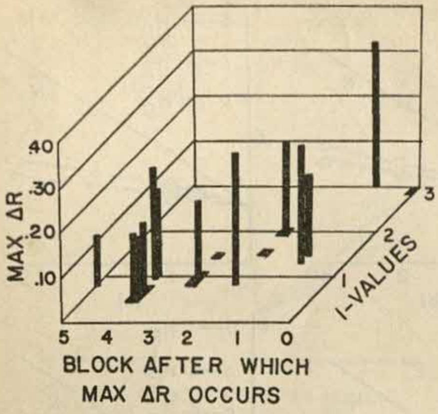


FIG. 5. Amplitude of MaxΔR, and the relation between the block after which it occurs and the *I* value of the feedback information for Exp. II.

same error was usually greater. For the four-alternative response conditions, when I_1 values were used, the percentage error between predictions of proportionate decrease and the obtained results was 15.3 for Cond. Q and 16.4 for Cond. H; when I_2 values were used the same errors were 7.1% and 4.7%, respectively. The corresponding values for the eight-alternative response conditions for Cond. Q and H, in the same order as above, were 17.3, 36.4, 34.3, and 95.3%.

Shape of the learning curve.—The relationship between the place of occurrence and value of MaxΔR is presented in Fig. 5 with I_1 values used to describe feedback information. The organization of this figure is identical to that of Fig. 2. The following probabilities were obtained in analyzing the data described in Fig. 5:

$$P_0 < \frac{1}{6^5} < .001$$

$$P_1 < \frac{1}{5^5} < .001$$

$$P_{Diag.} < \frac{1}{6^9} < .001$$

The results concerning the location of MaxΔR in the learning sequence are essentially the same as in Exp. I. As *I* values decrease, MaxΔR tends to occur later in the learning sequence. Thus, here also the shape of the performance curves vary with the *I* values.

The relationship between the *I* values and the value of MaxΔR was also analyzed by tau. When I_1 values were used the correlation was .515; for I_2 values, the correlation fell to .273.

EXPERIMENT III

The results from the immediately preceding experiment indicated that when the sequential pattern of noise deviated from a random pattern, the relationship between initial learning rate and the *I* values of the feedback was affected. In general less proportionality was evident. Experiment III is an investigation of the effects of various nonrectangular *random* distributions of noise lamp positions. It examines the prediction that the inverse proportionality between initial learning rate and amount of noise will be invariant over variations in the proportion of times a noise lamp is associated with a single position. In Exp. II the temporal distribution over trials of the noise lamps was altered; in Exp. III the spatial distribution of noise lamps is varied.

Method

Conditions.—This experiment consisted of 15 conditions, 2 of which were free of noise. The following are the conditions and their *I* values:

1. (12, 0, 0, 0), $I = 2.00$
2. (12, 0, 0, 0, 0, 0, 0, 0), $I = 3.00$
3. (12, 4, 4, 4), $I = .86$
4. (12, 4, 4, 4, 0, 0, 0, 0), $I = 1.29$
5. (12, 2, 2, 2, 2, 2, 2, 0), $I = 1.36$
6. (12, 6, 6, 0), $I = .67$
7. (12, 6, 6, 0, 0, 0, 0, 0), $I = 1.00$
8. (12, 6, 3, 3), $I = .75$

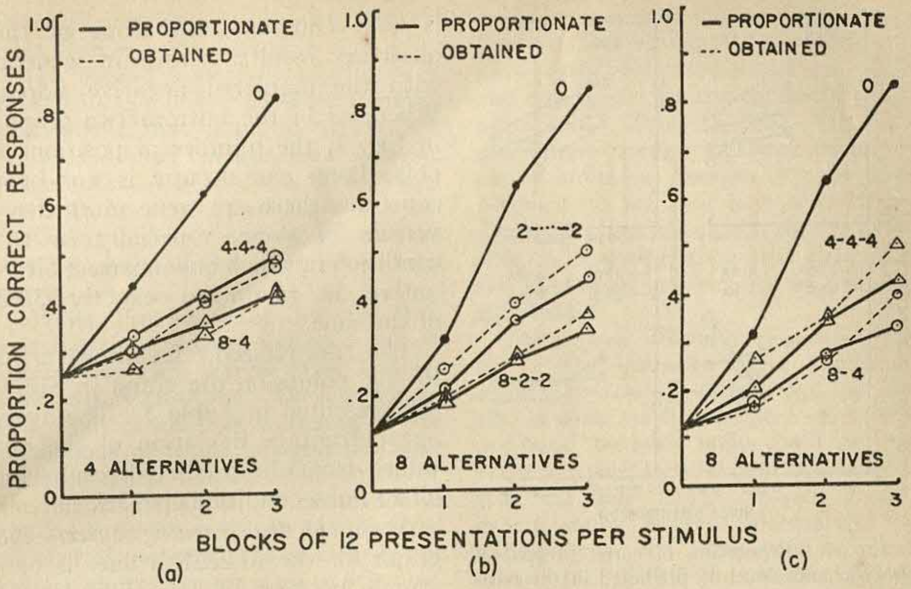


FIG. 6. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials for those conditions of Exp. III for which the differences between the proportionate and observed were least. (Part a contains four-alternative conditions, while Parts b and c contain eight-alternative conditions.)

9. (12, 6, 3, 3, 0, 0, 0, 0), $I = 1.12$
10. (12, 6, 1, 1, 1, 1, 1, 1), $I = 1.20$
11. (12, 8, 4, 0), $I = .58$
12. (12, 8, 4, 0, 0, 0, 0, 0), $I = .87$
13. (12, 8, 2, 2), $I = .59$
14. (12, 8, 2, 2, 0, 0, 0, 0), $I = .88$
15. (12, 8, 1, 1, 1, 1, 0, 0), $I = .90$

To refer to conditions briefly, a shortened notation listing only noise lights will be used, e.g., 4-4-4 or 8-1-1-1-1. Conditions with nearly rectangular distributions of noise lights such as (12, 2, 2, 2, 2, 2, 2, 0) will be abbreviated further, e.g., 2- . . . -2. A "0" will be used to refer to condition with no noise.

Subjects and stimuli.—Ten *S*s served in each condition. One of these *S*s was a replacement for an *S* who apparently failed to understand the instructions. The same criteria for rejection were used as in Exp. I. The eight stimuli were the same as those used in Exp. II.

Results

Initial rate of learning.—In Fig. 6 are presented the curves for those six noise conditions which had the smallest normalized deviation areas and the two noise free conditions. Part a gives results for conditions with four response alternatives; Parts b and c for those with eight response alternatives. The normalized deviation areas for these conditions are given in Table 3. The average percentage error in the prediction of proportionality for these six conditions is 10.6. It can be observed that the normalized deviation areas were smallest when the noise was either distributed rectangularly or when the noise was heavily concentrated on a single position.

TABLE 3

NORMALIZED DEVIATION AREAS FOR EXP. III

Noise Distributions	4-Alternative Conditions	8-Alternative Conditions
2-2-2-2-2-2	—	2.45[a]
4-4-4	1.39[a]	2.50[a]
6-1-1-1-1-1-1	—	3.33
6-3-3	3.24	6.33
6-6	4.42	3.52
8-1-1-1-1	—	2.87
8-2-2	5.51	1.70[a]
8-4	2.34[a]	1.45[a]

[a] These conditions are presented in Fig. 6.

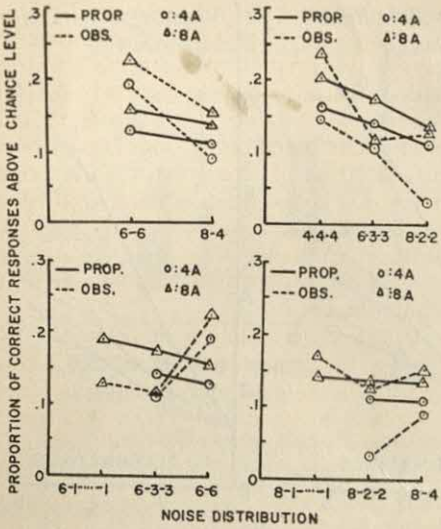


FIG. 7. Proportion of correct responses above chance level as predicted in the basis of proportionality and as observed for the conditions of Exp. III for the average of the first three blocks of trials. (Four-alternative conditions are symbolized by having data points included in circles while triangles are used for eight-alternative conditions.)

In Fig. 7 are presented the average predicted and average observed proportion of correct responses above chance level for the first three blocks of trials. To obtain these graphs, the proportion of correct responses which could be expected if *S* were responding randomly (chance level) was subtracted from the proportion of correct responses for each of the first three blocks of trials. This gave the proportion of correct responses above chance level for each block of trials. These results were then averaged over the first three blocks of trials. The points on the abscissa in Fig. 7 represent conditions. The conditions were arranged so that the predicted proportionate response values would always have a negative slope.

In the upper two graphs of Fig. 7 it can be seen that when the number of positions a noise lamp can occupy is kept constant, the slope of the observed results agrees in general with the predicted negative slopes. When, as in the bottom two graphs in Fig. 7, the number of positions a noise lamp can occupy is not kept constant, there are some marked reversals in slope, especially in the condition in which one incorrect alternative is reinforced exactly 50% of the time.

The normalized deviation areas for the points on the graph in Fig. 7 are presented in Table 3. The average percentage deviation of the obtained from the predicted results for all 13 noise conditions is 15.9.

Shape of the learning curve.—No graph for the MaxΔR values is presented because of the difficulty of presenting 15 conditions clearly. The results concerning the location of MaxΔR in the learning sequence are essentially the same as those in Exp. I and II:

$$P_0 < \frac{1}{6^8} < .001$$

$$P_1 < \frac{1}{5^5} < .001$$

$$P_{Diag.} < \frac{1}{6^{11}} < .001$$

The tau correlation between *I* values and values of MaxΔR was .543.

DISCUSSION

The most important conclusion to be drawn from these experiments is that it is possible to use a measure of feedback information *not* based on performance which permits fairly accurate predictions of initial performance on a learning task from a knowledge of only (*a*) performance in the task with no reduction in feedback information and (*b*) the amount by which the feedback information is reduced. Such predictions are not always very good, and indeed at times may be quite poor, but on the average they offer an initially acceptable basis for making quantitative predictions.

The second important conclusion is that the value of the feedback information affects that point at which performance shows its maximum increase in correct responses. Thus, there clearly seems to be some relationship between the value of the feedback information and the shape of the averaged learning curve.

While the preceding are the two principle conclusions of this set of experiments, there are several other findings which deserve mention. The first is more of a qualification than a finding. Since predictions of initial performance were independent of number of alternatives, the important part of the calculation in the prediction rested upon the expected number of trials required to identify the correct alternative for a given stimulus. Thus, the information measures used in making predictions for a fixed number of alternatives were directly dependent upon the expected number of trials over which information had to be retained.

Secondly the proportionality prediction lead to different errors in the case of four-alternative and eight-alternative conditions. In almost all cases, predicting a proportionate reduction for the eight-alternative conditions resulted in underestimating the obtained results. In many of the four-alternative conditions, a decrease in performance proportionate to the decrease in information fed back following a response would result in an overestimate of the obtained results. Thus, it is possible that the proportionate predictions might have been best for conditions employing approximately six alternative responses. Certainly, in the further developments of any model, the number of alternative responses must be considered as an important variable affecting the performance results of the reduced information.

With regard to these differences between four- and eight-alternative conditions it is also interesting to note the relative accuracy of the predictions based upon two different types of information measures in Exp. II. There were two conditions, Q and H, for which the I_1 and I_2 values gave different predictions. For the four-alternative conditions, I_2 gave better predictions than the I_1 values and conversely for the eight-alternative conditions, (Table 2, Fig. 3). The predictions based on I_2 values essentially placed stress on changes in noise lamp position. If the noise is more confusing when there are fewer alternatives available, it is possible then that changes in the noise lamp position assume a greater significance.

The greater difficulty experienced by *S*s in the four-alternative conditions cannot be related entirely to the fact that in some conditions there were fewer positions the noise lamp could assume. Were this true, it could have been argued that the noise lamps created greater generalization between alternatives by appearing in the same position more often. In Exp. III the number of positions the noise lamp could assume and the number of times it assumed each of these positions were controlled so that any generalization which occurred between lamps should have been equal for both the four- and eight-alternative conditions. Yet, as was observed in Table 3, even under these conditions, *S*s in the four-alternative conditions found the learning to be more difficult. There are two possible explanations. The first is that the simple availability of more positions the noise lamp could assume in the eight-alternative conditions, even though it never assumed more positions than in the four-alternative conditions, made their discrimination simpler. The other is that the need for *S* to associate *two* switches with each stimulus in the four-alternative conditions made the noise more effective in interfering with learning in these conditions.

There were also, on the other hand, certain similarities between the effects of the noise on performance for both four- and eight-alternative conditions. The most noticeable one is the U-shaped function which appears in Table 3. A prediction of proportionality was best when the distribution of the place of occurrence of the noise was either very close to a rectangular distribution, or highly concentrated in a single position.

The differences found between Cond. Random$_1$ and Random$_2$ in Exp. II may be partly accounted for in this way. Such a U-shaped function seems reasonable because an even distribution of the place of occurrence of a noise lamp makes the constant lamp more discriminable, while a concentration on a single position increases the discriminability of that single incorrect position for the remainder of the incorrect positions. This is in accord with Morin's (1955) finding that although awareness of misinformative feedback increased persistence in specific incorrect modes of response it also increased learning rate.

In these experiments again, a reinforcement of 50% produced some deviations from what could be described as the "overall pattern." In Exp. III, the largest deviations from proportionality occurred when one incorrect response was reinforced exactly 50% of the time. When only one alternative was reinforced exactly 50% of the time, there was always, for both the four- and eight-alternative conditions, less than proportionate use made of the feedback information. Reinforcing an incorrect alternative 50% of the time apparently attracted more responses away from the constant lamp than did, for example, a reinforcement of an incorrect alternative 66.7% of the time. (This may account in part for the previously described U-shaped function.) A similar result occurred in a study of the effects of increments of reinforcement (Shelly, 1960). There it was shown that the greatest decrease in responding to an alternative which was reinforced 75% of the time occurred when the probability of reinforcement on another alternative reached 50%.

SUMMARY

Three experiments are reported, each aimed at providing a partial answer to the following two questions: (*a*) When the information fed back to *S* concerning which of four (or eight) alternatives is correct, is reduced, is the initial rate of learning reduced proportionately? (*b*) Will the point at which maximum increase in proportion of correct responses takes place vary with the amount of information available to *S*, i.e., will the "shape" of the learning curve vary with amount of information fed back?

Three experiments, based upon a paired-associate paradigm, were conducted. Experiment I varied the amount of information, or conversely the amount of noise, in the feedback given *S* through variations in the number of "noisy events" which could follow a response; Exp II varied the sequential pattern of the noise; and Exp. III varied the distribution of the noisy events over the set of available incorrect alternatives.

The conclusions were that a prediction that noise reduces initial learning rate proportionately to the amount of noise can serve as a first approximation, and that there are differences between conditions in the point at which the maximum increase in proportion of correct responses occurs. The prediction of proportionality in initial learning rate was most accurate when there was a single noise lamp which was distributed evenly over the set of incorrect alternative positions, or when the noise lamp appeared very often in a single position. Consistent differences were noted between the effects of the noise in four- and eight-alternative conditions. It was more effective in depressing performance in the four-alternative conditions. For the range of conditions used, the following was generally found concerning variations in the point of maximal increase in correct responses: When such points of maximal increase were compared with feedback information, the relationship was quite close to linear.

REFERENCES

BOURNE, L. E., & PENDLETON, R. B. Concept identification as a function of completeness and probability of information feedback. *J. exp. Psychol.*, 1958, **56**, 413–420.

MORIN, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49**, 343–351.

NEIMARK, E. D. Effects of type of reinforcement and number of alternatives in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–228.

SHELLY, M. W. Patterns of decrease in response probabilities. *J. exp. Psychol.*, 1960, **59**, 345–350.

TROWBRIDGE, M. H., & CASON, H. An experimental study of Thorndike's theory of learning. *J. gen. Psychol.*, 1932, **7**, 245–258.

(Received April 14, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 223–233

BALL THROWING RESPONSES TO PHOTOGRAPHICALLY PORTRAYED TARGETS[1]

PATRICIA CAIN SMITH AND OLIN W. SMITH

Cornell University

This paper reports four experiments which were performed to determine how well photographs would serve as the distal stimuli for motor responses, e.g., tossing of balls to absolute distances as presented monocularly in the space immediately adjacent to *S*. The experiments are intended to provide an inductive base for an important but neglected class of perceptual problems: The determination of conditions of observation and visual stimuli which are conjointly necessary and/or sufficient for perceptions of absolute distances in the everyday world of *S*. Present theories (Gibson, 1950; Hebb, 1949; Ittelson, 1961; Koffka, 1935; Ogle, 1950) simply ignore this class of problems and in so doing they ignore problems of direct relevance to the everyday life of *S*. This unfortunate fact can be attributed to three causes.

The first is the general failure of theorists to regard *S*'s motor functions as determinants of many of the units by which *S* perceives nearby absolute distances in his everyday life. The assumption which we make here (for which there is abundant evidence) is that the space which is of greatest importance to the mobile *S* on a ground surface is the space of his immediate environment. It is in this space that *S* walks, jumps, and finds his daily bread. It is in this space that the most common instances of perceptions of absolute distances take place. Most of these perceptions involve motor acts. For instance, a puddle is perceived as too large to step across; a ditch is perceived as too broad to jump; a vacant bit of curbing is seen as too short in which to park a car.

The units of perceived absolute distance vary according to the motor acts by which *S* copes with his immediate spatial environment. The units are standard only in nomenclature. They are highly specific to each *S*; they vary within *S*s over time; they differ from *S* to *S* at any given moment. Yet, these are the units of perceived absolute distance which are most practiced in everyday life and most necessary for self-preservation. The units do not lend themselves readily to study, particularly under conditions permitting *E* the option of controlling the conditions of observation as well as selecting and varying the visual stimuli for responses. This difficulty alone may be the sufficient explanation for the general neglect of theory and research in this area.

A second cause for the failure of theory to account for absolute judgments is that light to the eye has been most frequently taken as the independent variable of visual perception (see Zajac, 1959, who would restrict visual perception to the study of perceptions which can be accounted for by ocular functions alone). Let us emphasize here that all other things equal, *S*'s posture and the position of his eye in space over time control the geometrical perspectivities of light which can appear on the retina in one-to-one correspondence with the distal stimuli of everyday life. The light entering the eye of *S* is dependent, then, upon the independent variables of *S*'s posture and motion in space (assuming that the *S*'s head is rarely static in everyday life). The *S* that is supported by a ground surface is usually aware of his absolute changes in spatial position independently of vision or concomitantly with the visual stimuli which confirm this change in position. Visual perception as studied traditionally, however, takes the visual stimuli as the independent variables. In order for the theory of visual space perception to be optimally relevant to the perceived absolute distances which are crucial for everyday life, theory must take into account the joint changes in *S*'s position in space along with the dependent changes in his visual stimuli.

---

[1] This research was supported in part by Contract AF18 (600-443) between Cornell University and Human Resources Research Laboratory, Bolling Air Force Base, Washington, D. C., and Contract Nonr 401(14) between Cornell University and the Office of Naval Research. The assistance of P. Herman and A. Steinschneider is gratefully acknowledged.

A third cause has been the absence of any suitable apparatus permitting: (*a*) the selection, control, and variation of visual stimuli in conjunction with the postures which are customary (or nearly so) for the distances to be viewed; (*b*) postures appropriate to these distances. For example, the apparatus for exposing the visual stimuli for the space near the feet of *S* should require an appropriate tilt of *S*'s head for the purpose of aiming his eye at something that he sees.

The apparatus for satisfying these conditions for binocular vision in the mobile *S* might be difficult to achieve. A simple and inexpensive apparatus for monocular and static views, however, seemed possible. The research on photographically portrayed spaces suggested the use of suitably mounted photographs as the distal stimuli for motor responses corresponding to perceived absolute distances in the portrayed spaces. Monocularly viewed photographs would also permit (*a*) the study of visual direction as a stimulus for perceived absolute distances, and (*b*) the assessment of perceptions of absolute distances that are independent of the cues of accommodation and convergence. These latter cues could only carry information about the optical distances of the photographs, not of the portrayed distances; any effects attributable to accommodation and convergence would be interference effects. A test of the effectiveness of such an apparatus for motor responses to distances perceived under "optimal" visual stimulation in photographs would serve as a criterion for evaluating the extent to which other combinations of stimuli would be sufficient for perceived absolute distances.

Schlosberg (1941) demonstrated that photographs can induce compelling and highly realistic impressions of depth when *S*'s monocular field of view is restricted to a photograph. The heightened realism of this effect is probably due to the fact that when *S*'s field of monocular view is not restricted to the photograph, *S* perceives two spaces, the space portrayed in the photograph and the three-dimensional space immediately adjacent to *S* of which the photograph as an object is only a part. Under the normal conditions in which photographs are viewed, the two spaces, that is, the portrayed space and the space of *S*, are incompatible. Restriction of the field removes this incompatibility.

Photographs can also subtend to the eye visual angles which closely approximate those subtended by the objects themselves to the lens of the camera (Henney & Dudley, 1951, pp. 104–106). Gibson (1954) assumed that if photographs were viewed appropriately, the resultant perceptions of distance should be equivalent to those induced by the original scenes. This assumption was tested and found tenable (but subject to constant errors) by Smith and Gruber (1958) who compared directly the distances perceived in a still photograph with those perceived in the original space. We therefore assumed that when other factors were held constant, photographs of targets in the space immediately adjacent to *S* could be substituted for direct views of the targets themselves.

The distance to which a ball is tossed at a photographically portrayed target is an excellent distance response with which to undertake the investigation of this neglected area since (*a*) the extent of the toss is an absolute measure of the distance response, (*b*) most male adult *S*s are well practiced in making this response to nearby targets in their immediate environment, (*c*) the response can be made to the different distances of *S*'s immediate environment, and (*d*) the response can be made under restricted as well as unrestricted conditions of viewing.

The use of photographs, however, as the distal stimuli for targets of tossed balls involves variables in addition to those of perceived distances. For instance, the degree of restriction of *S*'s monocular field of view is important in that a sufficiently large field of view can induce the impression of a space which is immediately adjacent to *S*'s person (Smith, 1958). Conversely, a minute field of view can create the impression that the viewed scene is detached and not a part of *S*'s immediate environment. Consequently, three different sizes of field of view were studied.

Another variable is the visual direction of a target. Bailey (1954) demonstrated that the visual direction of an elevated target can be given in perception by a single point of light in an otherwise dark environment. Hence, the traditional cues for depth and distance of a target relative to *S* are not necessary for establishing its direction. A different case is that of the visual direction of a target resting on a ground surface that is the basis of *S*'s support. In this instance, the absolute distance of the target from *S* is also a determinant of the visual direction of the target. Photographs displayed so as to cause *S* to tilt his head also establish for him the visual direction of the target as portrayed on a ground surface common to *S* and the target. Are the conditions of establishing the visual direction of targets in photographs sufficient for the perception of the absolute distances of the targets? This question was therefore investigated both as a control for the picture experiment and as a problem in its own right.

EXPERIMENT I

This experiment investigates the effects of introduction of three sizes of apertures in a monocular field of view on the accuracy of tosses of balls at directly viewed targets.

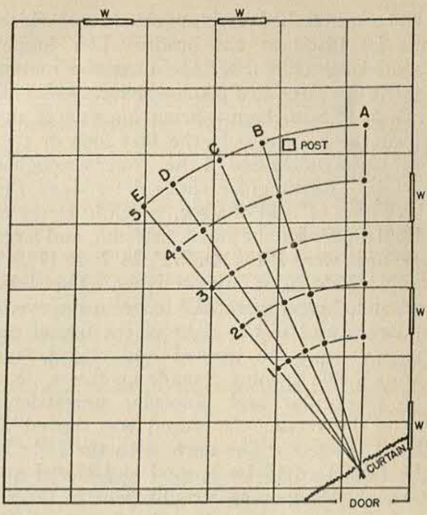


FIG. 1. Diagram of room of the experiments. (Dots indicate the target positions.)

Method

The experimental room was 36 × 30 ft., and illuminated by fluorescent tubes in industrial fixtures. The floor was made of concrete blocks 5 ft. square and was dusted lightly to make clearly visible the point of impact of the tossed balls. There were 25 target locations at five distances: 3, 4.25, 5.5, 6.75, and 8 meters from a point directly under the eye of the viewing apparatus. These will be called Distances 1 through 5, respectively (see Fig. 1). For each of the distances, five stations were located at different angles from the viewing position. A flat yellow target, 8 in. in diameter, was placed on the station specified for each trial.

When *S* entered the room (see Fig. 1) he saw only a curtain of black cloth and an eye piece which extended through a slit in the cloth. Behind the cloth was a rectangular wooden stand which was 72 in. long, 45 in. high, and 12 in. broad. The stand was located approximately parallel to and in the middle of the distance allocated to the curtain in Fig. 1. In the middle of the top plank of the stand was mounted the head of a tripod. The head of this tripod could be swivelled in either or both the horizontal and the vertical and locked into position. To this head was attached a rectangular funnel which was jet black on the inside. This funnel tapered from 2 in. × 1$\frac{21}{32}$ in. at the mount for the eye piece to a plane which was 18.5 in. wide × 15.5 in. high. From *S*'s side of the apparatus he saw only the first inch or two of the protruding end of the viewer through a curtain. Rectangular slides determined the three sizes of field of view, which in terms of visual angles for the small, medium, and large apertures were 7°12′ × 5°12′, 24°7′ × 17°54′, and 41°7′ × 30°38′, respectively. The slides in a slide holder were 20.8 in. from the eye.

Below and to the right of the funnel the curtain could be opened and closed, and through this opening *S* made his tosses. For the monocular and binocular unrestricted series the rectangular stand was moved 20 in. to the left. The curtain to the right of the funnel could be opened and closed and through this opening *S* could view the targets with unrestricted head motion for controlled periods of time. As before, he tossed his balls through a controlled opening on the lower side. A regulation soft ball weighing 6½ oz. was used. The target was centered in the field of the viewing apparatus for all aperture trials. The *S*s did not observe this process.

The *S*s for Exp. I were 50 inductees from Sampson AFB, New York. Each was given practice and instruction in tossing a ball at a target underhanded, in a space different from that of the experiment. He was made familiar with positioning his eye at the eye piece, and was given practice in tossing as hard as he could, as well as gently, through the opening in the stand without viewing a target for his tosses while his eye was in position at the eye piece. Some practice in tossing was necessary so as to insure that *S* would release the ball before the upsweep of his hand reached the top plank of the stand on which the tripod was mounted. He was practiced in closing his eyes when the ball left his hand. To insure the absence of visual knowledge of results of his throws, *E* barred *S*'s field of view with a flag as soon as *S* tossed a ball. No verbal information was ever given *S* regarding the accuracy of his tosses. A count-down system was used to insure nearly equal times of exposure of all targets for all views through the viewing apparatus as well as during monocular and binocular unrestricted views. The *S* was instructed to toss at the count of four. Some head motion parallax was possible under all conditions during the short period in which the targets were viewed.

All *S*s made one toss at one of the five target locations at each of the five distances for small, medium, and large apertures as well as for monocular and binocular unrestricted views in that order. The order of distance and location of target for any one viewing condition was randomized for each *S*. The only restriction was that all five distances were presented under each viewing condition. One *E* instructed *S*, adjusted the viewer, and operated the flag while two *E*s retrieved the balls, measured the distances tossed, recorded, and positioned targets. At no time did *S* see the *E*s beyond the viewing stand, or the room, between throws during the restricted series.

The zero point from which measurements of distance originated was a point directly beneath the eyepiece of the funnel. When the stand was displaced 20 in. to the left for the unrestricted series, the origin of measurement was not displaced. Lateral errors of tosses were not measured since in preliminary trials these were a function of the distance of the targets and exhibited a narrow range of dispersion when compared with the dispersion of the tosses in distance.

TABLE 1

ANALYSIS OF VARIANCE, EXP. I

Source of Variation	*df*	*MS*	*F*
Condition of viewing (C)	4	916.37	2.77
Distance (D)	4	11,300.97	72.48
*S*s	49	2,499.38	55.86
C × *S*s	196	331.29	7.40*
D × *S*s	196	155.92	3.48*
C × D	16	169.79	3.79*
C × D × *S*s	784	44.74	
Total	1249		

* $P < .01$.

Results

All *S*s were able to toss the ball at the targets. Under aperture conditions 23 *S*s hit one or more targets for a total of 37 hits. A number of *S*s inquired about the (nonexistent) lens system or glass in the viewing apparatus.

The distance of the point of impact of the ball on the floor was converted to a percentage of the target distance. These numbers were subjected to an analysis of variance which is sum-

TABLE 2

MEANS AND *SD*s OF PERCENTAGES OF TARGET DISTANCE THROWN FOR VIEWING CONDITIONS OF THE FOUR EXPERIMENTS

Exp.	Conditions of Viewing Apertures								Unrestricted Conditions			
	Small		Medium		Large		Combined		Monocular		Binocular	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
I	100.1	12.8	99.5	15.1	99.8	13.1	99.8	12.0	96.1	10.1	96.6	9.0
II	94.4	9.7	96.3	9.2	102.4	10.7	97.7	9.1	92.4	8.6	91.7	6.7
III	103.0	23.3	104.2	25.4	106.1	26.8	104.4	22.9	101.1	9.7	97.3	7.4
IV	87.2	21.4	86.7	27.8	85.5	24.8	86.4	23.4				

marized in Table 1. The second order interactions were all significant at the 1% level. The means of all distances for the small, medium, and large aperture conditions are nearly the same and are larger than those for the two unrestricted conditions (see Table 2). According to a method of analysis suggested by Tukey (1949) for the determination of sources of variation when certain treatments are known to be different as a class from certain other treatments, most of the variation associated with conditions of viewing is attributable to the difference between aperture and non-aperture conditions (see Table 3).

Target distance was a major source of variance despite the conversion of raw scores into percentage of target distance. In order to determine which of the means were significantly different from the overall mean for distances, Tukey's (1949) method was used for breaking up treatments into groups where there is no a priori basis for grouping and where an analysis of variance shows that not all treatments are alike. By this analysis the mean for Distance 5 is the only one which is significantly different from the overall mean. The far target may have been consistently undershot because it was the end point of the series.

Discussion

Restricting the field of view of naive *S*s to the surround of their targets had the overall effect of causing them to throw harder (as if the targets appeared more distant) than when views of the

TABLE 3

MEANS AND *SD*s OF PERCENTAGES OF TARGET DISTANCE THROWN OVER ALL APERTURES IN THE FOUR EXPERIMENTS

Exp.	Distance of Target in Meters										Total	
	3		4.25		5.50		6.75		8.0			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
I	107.0	15.8	105.7	14.6	101.0	14.5	95.6	12.7	89.8	12.4	99.8	12.0
II	103.9	13.2	103.6	8.9	99.2	10.7	92.3	10.9	89.4	9.0	97.7	9.0
III	114.2	15.2	109.5	19.6	104.9	14.7	99.6	14.4	93.9	10.5	104.4	22.9
IV	101.9	33.4	92.9	22.6	87.8	29.7	79.0	21.6	70.7	16.4	86.4	23.4

targets were not restricted. Had *S*s been permitted to view the points of impact of their tosses and to correct for errors it seems very likely that the difference would have been eliminated.

The Conditions × Distance interaction was predicted in advance of the experiment since use of the motor response as a criterion for perceived absolute distance as well as the restriction of the fields of view introduced unavoidable interactions. For instance, when the angular extent of a restricted field of view is held constant for views of targets at different distances, the visible distances on the ground of the target areas vary with the distance of the target. The ratio of the distance viewed within a target area to the distance separating *S* and the nearest visible portion of a target area decreases and then increases as a function of the distance of the target (assuming a level ground common to *S* and target).

The motor function of positioning the head for inspecting a target should have its greatest cue value for distance for the target nearest to *S* since the greatest inclination of the head is required. Also, for any one field of view the ratio of the distance of surface viewed in the target area to the distance separating *S* and the nearest visible point of the target area is greatest at near position. And, lastly, in the absence of knowledge of errors of response, the distances to which balls are tossed should decrease with an increase in target distance beyond a neutral point. These comments only serve to indicate that all other factors do not remain constant when the distance of targets is increased.

EXPERIMENT II

This experiment was performed to determine whether the order of observations in Exp. I had a significant effect.

Method

The serial order of conditions of observation was reversed from those of Exp. I; i.e., the first condition of tossing was with unrestricted binocular views which gave *S* a complete view of the area in which the targets were spotted at different locations and distances. Instructions and practice were essentially the same for the two experiments. The *S*s were 16 inductees from Sampson Air Force Base.

Results and Discussion

Twelve of 16 *S*s made a total of 19 target hits in Exp. II under aperture conditions as compared with only 23 of 50 *S*s who scored a total of 37 hits in Exp. I. This difference in target hits might be attributed to either or both of two causes. The first is that the additional information gained from the binocular and monocular views carried over to performance under aperture conditions. The second interpretation that *S*s of this experiment were, as a group, more accurate ball tossers than those of Exp. I, is not tenable since neither the means nor variances for monocular unrestricted views of the two experiwere significantly different. The first possibility remains tenable.

The directions of the overall differences between aperture and unrestricted views are the same in the two experiments. An effect attributable to serial order between means of the small and larger apertures was significant by *t* test ($P < .01$). Effects of target distance are also parallel to those of Exp. I (see Tables 2 and 3).

EXPERIMENT III

This experiment repeated Exp. I with the exception that *S*s observed photographs in the viewing apparatus instead of the room itself. The purpose of this experiment was to compare the accuracy of responses to photographs with those of the criterion group in Exp. I.

Method

Photographs were made of all targets as viewed under the large aperture condition of

Exp. I. The lens of the camera, a Graphic View II, was placed at the eye position of *S* in the viewing apparatus. The film was panchromatic X; the prints were on K-2 paper; and the lens a 152 mm. Ektar. Uniform conditions were maintained in developing and printing. The prints were achromatic with the exception of the target which was tinted a yellow to match the color of the original target. This permitted easy identification of the targets and use of instructions nearly identical to those of Exp. I. There was no attempt to achieve high verisimilitude, but rather accuracy of angular relationships between print and scene, and comparability of conditions of viewing and instruction.

Each print was bound between two glass plates. When a slide was inserted in the slide holder of the viewing apparatus, the angles subtended to the eye of *S* by objects in the photograph corresponded exactly to the angles subtended by the original objects. The slides were backlighted by the ordinary room illumination. This backlighting caused the photographs to be darker than the real scene. The high contrast between the level of brightness of the picture and the black walls of the viewing apparatus compensated perceptually for the lower brightness level of the pictures. The viewing apparatus was adjusted for direction and elevation so that it was centered on each target before the picture of that target was inserted. The *S*s did not witness this positioning of the viewing apparatus.

The *S*s were 30 inductees from Sampson Air Force Base. The instructions for this experiment are those of Exp. I, with the exception that after each *S* had completed all of his tosses he was shown the slides in their rack and was asked if he recognized them. He was also asked if there was anything unusual about what he had seen through the viewing apparatus.

Results and Discussion

No *S* reported that he had seen photographs in the viewing apparatus, or noted any unusual quality of what he had seen, even its achromatic nature. In this respect, the photographs were unexpectedly successful as surrogates for the room and its targets. All *S*s recognized that the slides were pictures of the room, without understanding the mode of their use.

Their accuracy was also surprisingly good. Thirteen *S*s made a total of 14 target hits while viewing photographs. The variance of the combined aperture responses to pictures was significantly greater ($P < .01$) than the corresponding variance of Exp. I. This difference in variance could account for the proportionately fewer hits made by the picture group than by the groups of Exp. I and II.

The mean for each of the three apertures in which the pictures were viewed is higher than the corresponding mean of Exp. I. The mean and variance of the *monocular* unrestricted responses are assumed to be valid measures for assessing "basic" differences in ability between the groups of Exp. I and III independently of responses under aperture conditions. The variance of the two groups did not differ significantly, but their means did differ ($P < .05$). Therefore, the means of the responses to pictures should be corrected by the amount of the difference between the means of the monocular unrestricted responses before the means of the responses to pictures can be validly compared with responses to direct views of the targets under aperture conditions. The subtraction of five percentage points from the mean for responses to photographs nearly eliminates the differences between aperture means for Exp. I and III. Therefore, the mean effects of photographs did not differ significantly from aperture effects or direct views of the targets. Direct views of the targets, however, did permit some monocular motion parallax which would have accounted for the significant reduction in variance.

Motion parallax and binocular disparity were effectively eliminated as cues or stimuli for perceived depth and distance by the use of photographs.

Accommodation and convergence were made for the optical distance of the photographs. If functional, their effects would have been to decrease the magnitude of the motor responses. The means (see Tables 2 and 3) provide no evidence for such effects. Motion parallax, binocular disparity, accommodation, and convergence were eliminated as cues for the absolute distances involved in the motor responses. The *SD* for the monocular unrestricted condition was only slightly greater than for the binocular unrestricted condition. This indicates that nearly all of the increase in accuracy in the unrestricted conditions is attributable to accommodation, convergence, and motion parallax in combination, and not to binocular disparity.

The *S*s of this experiment frequently reported that they did not feel sure of the results of their tosses. This feeling of uncertainty is undoubtedly due to the absence of the above four cues.

EXPERIMENT IV

This experiment determined whether the visual direction of a target on a surface which also supported *S* would serve as a stimulus for the distance of the target from *S*. The apparatus for the aperture and photograph experiments caused *S* to tilt his head as well as to adjust its left-right position. The distance to which a ball was tossed was again used as evidence of the effectiveness of the stimuli for distance.

Method

The same viewing apparatus, aperture conditions, targets, and target stations were used as in Exp. I, II, and III. Crosshairs were centered on a slide so that the position of the intersection of the crosshairs was in the center of the rectangular field of the viewing apparatus when the slide was inserted in the slide holder. A heavy sheet of white paper was mounted so that it could either cover the exterior opening of the viewing apparatus or reveal it. With this paper cover in place *S* could see only black crosshairs against a white field. Before each judgment by *S*, *E* adjusted the viewing apparatus so that the crosshairs were centered on a target. As in the other experiments the order of distance tossed, as well as the selection of target at any one distance, was randomized for each *S*. The instructions were essentially the same as for Exp. I except that *S* was told that a straight line from his eye through the crosshairs would intercept his target on the floor. The *S* was permitted no views of the room in which were located the targets. No information was given *S* regarding the accuracy of any response. The *S*s were 10 inductees from Sampson Air Force Base.

Results and Discussion

Some *S*s had difficulty in tossing balls at the points where they estimated the targets to be. Only one target hit was scored. The means for percentages of target distance thrown decrease progressively with increases in distance of the targets. This holds for Exp. I, II, and III (see Table 3) as well, but the rate of decrease in Exp. IV was accelerated greatly. The mean for Distance 1 more closely matched those of Exp. I, II, and III. In Exp. I the aiming function of positioning the head was hypothesized to be most effective for the nearest distance since the greatest inclination of the head from the horizontal is required for inspecting a target at 3 meters. The hypothesis remains tenable in terms of the obtained mean, but is false in terms of the *SD* which is largest at the near distance. The hypothesis requires further data for evidence of its tenability.

The evidence of this experiment is that the aiming function of the head in establishing the visual direction of targets located at 3 to 8 meters is sufficiently informative to *S* (for both direction and distance) to account for a substantial proportion of mean distance to which balls were tossed in Exp. I, II, and III. The range of distance responses was from 101.9% for the 3-meter target to 70.7% for the 8-meter target. An increase in the distance of the targets would probably have caused the mean percentage distance of response to shrink much more.

The positioning of the eye for inspecting a nearby target on the ground is a normal function of *S*'s posture. This posture need not be exact since the vertical and horizontal motions of the eye that supplement postural adjustment complete the process of establishing the visual direction of the target. As the distance of ground targets passes through the space immediately adjacent to *S* the necessity for changes in the tilt of the head diminishes since relatively small vertical rotations of the eye can then scan, without strain to *S*, targets located as far away as the horizon. Theoretically, the informative value of visual direction as a stimulus for distance would rapidly become zero as the line of regard approached the horizontal.

All of the *SD*s for this experiment were greater than the corresponding *SD*s for target distances in Exp. I, II, and III (see Table 3). Even though the responses to distances of the targets in this experiment were not due to chance, the *SD*s were large enough so that any statement about visual direction as a stimulus sufficient for everyday survival behaviors of *S* would be exceedingly optimistic.

The visual perceptions of distance required for motor responses to targets in the immediately adjacent space of *S* is clearly a complex process. Traditional studies of the perception of distance take into account only the depth cues or stimuli found in the light as projected on the retina. The principal finding of this experiment is that when all other factors are constant, the function of aiming the eye by causing the head to be positioned is a determinant of perceived distance (assuming a level ground common to both *S* and the nearby targets).

DISCUSSION

The effect of target distance was common to the four experiments. The *S*s undershot the far targets and overshot the near targets (3 meters). The mean effects over apertures of increasing target distance was a decrease in the mean percentage of target distance of the tossed balls. There were no reversals of the trend, which held also for the monocular and binocular unrestricted distance responses. The distances of the targets were all well within the physical capabilities of tossing of all *S*s, as demonstrated in their familiarization trials with the apparatus. The most probable cause for the tendency to undershoot far targets was the absence of visual or verbal knowledge of results for a task which would probably improve in accuracy with formal training.

Accommodation and convergence in the photographic series of Exp. III were made for the distance at which the photographs were presented while the concomitant perceptions were of the spaces and distances for which the photographs were surrogates. Accommodation and convergence were necessary for the perceptions, but the latter were otherwise independent of these ocular functions. The spatial significance of the light dominated the perceptions. A tenable inference is that accommodation and convergence are probably of no effect as cues for the perceived depths and absolute distances over surfaces which are characteristic of everyday life. This does not mean that accommodation and convergence cannot be demonstrated to be sufficient for perceptions of depth relationships of small objects or points. The relevance of such demonstrations to the spatial perceptions of everyday life remains to be empirically determined.

The only theory to which these data may be relevant is the S-R theory of Gibson (1950, 1954) who would account for perceptions of space and distance in terms of gradients of light in the optic array. Given the gradients of light as transmitted by the photographs in Exp. III, the Gibson theory would predict "veridical" perceptions of depth and distance. The data of these experiments indicate that Gibson's theory is both adequate and inadequate depending upon the criteria which are used to evaluate it. Perceptions of both depth and absolute distances were obtained in Exp. III, thus confirming Gibson's theory. On the other hand, the most important question is whether the theory is adequate to account for the visual perceptions of absolute distances in *S*'s immediate environment. At this point the answer must be no, since the *SD*s for the distance responses to the portrayed targets are nearly twice those of the aperture conditions in Exp. I. It is, of course, possible that practice with verbal correction under the conditions of Exp. III and IV would cause the *SD*s of the distance responses to equal or be less than those of the aperture conditions of Exp. I. Granting this possibility it should be evident that the ultimate theory should be one which will account for the "nature" that *E* samples as well as the "nature" that he can create.

The data of Exp. IV demonstrate that distance responses dependent upon visual direction can be made in the absence of gradients of light. For these data the Gibson theory is irrelevant and therefore not necessary. The introduction of the light from the photographs, however, increased the number of target hits as well as the mean percentages of target distances of the tosses. Certainly the visual world adds to the kinesthetic world information important to the ability to perform.

The changes in perspectivities of visual stimulation which are dependent upon the absolute changes in spatial position of *S*'s eye were immobilized by the fact that head movement had no such effect with the photographs of Exp. III. The static images of the photographs served to confirm the posture taken by the head in positioning the eye, but not the minor changes in position of the eye during observation. This degraded the visual stimuli which are usually available in everyday life even though the light to the eye was in one-to-one or nearly one-to-one correspondence with the distances of the targets on the floor. The decrease in accuracy of the distance responses resulted.

If changes in the light to the eye confirm changes in the absolute spatial position of *S*'s eye, how does the light acquire significance for absolute distance responses such as ball tosses? The visual knowledge of the results of a tossed ball involves perceptions in addition to those of absolute distance. For instance, the perceived location of the point of impact of a ball with respect to a target is not dependent upon the perceived absolute distance separating the point of impact and target, e.g., the impact may be perceived as "under" or "over." The perception of "under" may result in additional force being applied to the succeeding toss. Should the perceived location of the second toss satisfy *S*, the perceived force of that toss would then become the unit of perceiving the absolute distance of the target. According to this rationale, appropriate units of perception for absolute distances can evolve as a consequence of perceptions of change of relative location of objects brought about by changes in motor functions. Hence, changes in the perspectivities of light on the retina, other things equal, may confirm not only changes in the spatial location of *S*'s eye, but also changes in the relative location of objects as brought about by other functions of *S*'s body.

SUMMARY

The evidence from four experiments involving ball tossing demonstrates that: (*a*) Photographs, when viewed appropriately, can serve as the distal stimuli for the motor response of tossing a ball to the distances of targets in the space immediately adjacent to *S*. (*b*) Restriction of the field of view of the viewing apparatus increases the extent of the perceived distances as defined by the motor responses. (*c*) Accommodation, convergence, motion parallax, and binocular disparity in combination do not have to be in correspondence with the distances to which accurate motor responses (target hits) are made. (*d*) Visual direction alone is an insufficient basis for accurate motor responses to the distances of the targets. (*e*) Photographs may be a useful tool for controlling the light to the eye for the study of visual

stimuli for perceptions of absolute distances relative to *S*.

REFERENCES

BAILEY, N. J. Determination of the location of the center, or centers of projection in anomalous correspondence. Unpublished doctoral dissertation, Ohio State University, 1954.

GIBSON, J. J. *Perception of the visual world.* Boston: Houghton Mifflin, 1950.

GIBSON, J. J. A theory of pictorial perception. *Audiovis. commun. Rev.*, 1954, **2**, 3–23.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HENNEY, K., & DUDLEY, B. *Handbook of photography.* New York: McGraw-Hill, 1951.

ITTLESON, W. H. *Visual space perception.* New York: Springer, 1961.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

OGLE, K. N. *Binocular vision.* Philadelphia: Saunders, 1950.

SCHLOSBERG, H. Stereoscopic depth from single pictures. *Amer. J. Psychol.*, 1941, **54**, 601–605.

SMITH, O. W. Comparison of apparent depth in a photograph viewed from two distances. *Percept. mot. Skills*, 1958, **8**, 79–81.

SMITH, O. W., & GRUBER, H. Perception of depth in photographs. *Percept. mot. Skills*, 1958, **4**, 307–313.

TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.

ZAJAC, J. L. Depth perception of stereoscopic images resulting from fusion of crossed and uncrossed images. *Amer. J. Psychol.*, 1959, **72**, 163–183.

(Received May 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 234–242

EFFECTS OF AMOUNT AND PERCENTAGE OF REINFORCEMENT AND NUMBER OF ACQUISITION TRIALS ON CONDITIONING AND EXTINCTION[1]

ALLAN R. WAGNER[2]

State University of Iowa

The relative performance levels obtained under partial and continuous instrumental reward training have been characterized by Goodrich (1959), Haggard (1956), and Spence (1960, Ch. 6) as including: (*a*) during early acquisition trials, a decrement under partial reinforcement which is more pronounced or more persistent as performance is measured closer to the goal region; (*b*) during asymptotic acquisition trials, superiority of partial reinforcement on response measures relatively early in the response chain; and (*c*) superior extinction performance following partial reinforcement.

Of primary interest in the present investigation was a determination of the way in which magnitude of reward influences the two mentioned acquisition effects, and the way in which magnitude of reward and number of acquisition trials jointly influence the extinction effect. Such determinations were thought to be particularly relevant to those theoretical interpretations (e.g., Amsel, 1958; Beier, 1958; Logan, 1960; Spence, 1960, Ch. 6) of partial reinforcement phenomena which have stressed the importance of anticipatory reward ($r_g - s_g$) and anticipatory frustration ($r_f - s_f$), both of which are assumed to vary with reward magnitude and number of acquisition trials.

METHOD

Subjects

The *S*s were 64 male and 64 female experimentally naive hooded rats from the colony maintained by the Psychology Department of the State University of Iowa. Their ages ranged from 125 to 130 days at the beginning of the experiment.

Apparatus

The apparatus was a straight, enclosed pine runway consisting of a 15-in. entry box, a 12-in. startbox, a 33-in. alley, and a 15-in. goalbox. All sections of the runway were 3 in. wide with the exception of the startbox which was narrowed to 2 in. Opaque doors separated the adjacent sections but, with the exception of the start door, were closed only to prevent retracing. A clear glass ceiling covered all sections of the runway at a height of 4 in. The walls and floor of the entry and startboxes were painted a flat grey, those of the alley and goalbox a flat black.

The apparatus was located in a black, sound-deadened room. Dim, diffuse illumination was provided by the light reflected from the black walls and ceiling and originating from four shielded 150-w. fluorescent tubes placed 4 ft. above the runway. A large fan located directly beneath the runway provided a general masking noise.

The apparatus provided three time measures of locomotor performance in the runway. Starting time was measured from the opening of the start door to the interruption of an infrared beam located 6 in. distant. Running time was measured from the interruption of this first beam to the interruption of a second beam located 12 in. beyond the start door. Goalbox time was measured between the interruptions of infrared beams located 7 in. and 1 in., respectively, from the end of the runway. The last beam passed just above the front lip of a stationary, ¾ in. deep, metal food cup.

---

[1] This research was conducted during the the author's tenure as a National Science Foundation Predoctoral Fellow at the State University of Iowa, and comprises part of a dissertation submitted to the Graduate College of that University, in partial fulfillment of the requirements of the PhD degree. The author is indebted to Kenneth W. Spence for his advice and assistance throughout the course of the investigation.

[2] Now at Yale University.

Procedure

Experimental design.—The experimental design included two levels of three variables: percentage reinforcement (100% vs. 50%), magnitude of reward (.08 gm. vs. 1.0 gm.), and number of acquisition trials (16 vs. 60). The design may be schematized as a 2×2×2 factorial, 16 *S*s, 8 males and 8 females, being randomly assigned to each of the eight experimental groups.

Habituation.—Thirty days prior to the first training day *S*s were placed on a 24-hr. food deprivation schedule, with 30-min. access each day to an ample supply of wet mash prepared from ground Wayne Lab Blox. This maintenance schedule was employed throughout the course of the experiment, while water was continuously available except during the experimental sessions.

On the 2 days prior to experimental training *S*s were allowed to explore the runway in groups of three for 5-min. periods. After the exploration on each of these days *S*s were placed in individual grey goalboxes and allowed to eat a number of reward pellets from a food cup identical to that used in the runway.

Experimental training.—A training trial was initiated with the introduction of *S* into the entry box. When *S* had entered the narrow startbox the retrace door was lowered and 3 sec. later the start door was raised. (The 2-in. width of the startbox discouraged *S* from turning so that all *S*s remained oriented toward the start door during the 3-sec. interval.) When *S* had entered the goalbox the retrace door was closed and *S* was removed 20 sec. later on nonreinforced trials or immediately after eating on reinforced trials.

The *S*s were tested in squads of 8, one *S* from each of the experimental groups. Eight such squads of males were tested first on each experimental day followed by eight squads of females. The *S*s were tested in the same sequence in every experimental session, but the order of testing was systematically varied among the squads so that equal numbers of *S*s from each group were assigned to each of the squad positions. Following the testing of all 8 *S*s in a squad, they were returned to their home cages, and 30 min. later *S*s were given their daily feeding. According to this schedule *S*s were tested under approximately 22½–23 hr. food deprivation.

Acquisition.—Half of the *S*s received 16 and the other half 60 acquisition trials, 1 trial per day, during which the percentage reinforcement and magnitude of reward were manipulated. The sequence of reinforcement and nonreinforcement for the partial reinforcement groups was arranged as follows: +−+− −++− −++− −−++ +−−+. It should be noted that this sequence has the following characteristics: (*a*) two reinforced and two nonreinforced trials appear in each successive block of four trials; (*b*) the first and last trials are reinforced; (*c*) never more than three reinforcements or nonreinforcements occur in a row; (*d*) when this 20 trial sequence is repeated the first order sequential probabilities are counterbalanced, i.e., reinforcement equally as often follows a reinforcement as it does a nonreinforcement.

The reward pellets were pressed from wet mash prepared immediately preceding the running of each squad of *S*s. Two pelletmakers were used which were designed to form pellets of .08 gm. and 1.0 gm. dry weight, when the mash was prepared by mixing .5 ml. distilled water with each gram of ground Wayne Lab Blox.

Extinction.—Extinction Trial 0 was administered to each group the day following its last acquisition trial. Extinction Trials 1 to 32 were then run on consecutive days. On each extinction trial *S* remained in the nonreinforced goalbox for 20 sec.

RESULTS

Acquisition

Mean starting, running, and goalbox speeds over blocks of four acquisition trials are plotted in Fig. 1 for the treatment groups receiving the four combinations of percentage reinforcement and magnitude of reward. Measures from all *S*s, i.e., 16- and 60-trial groups combined, are included in the curves over the first four blocks of acquisition trials, whereas the subsequent blocks of trials represent means from only those *S*s in the 60-trial groups. Over the first 16 trials, in all three response measures, large reward produced faster speeds than did small reward, while continuous

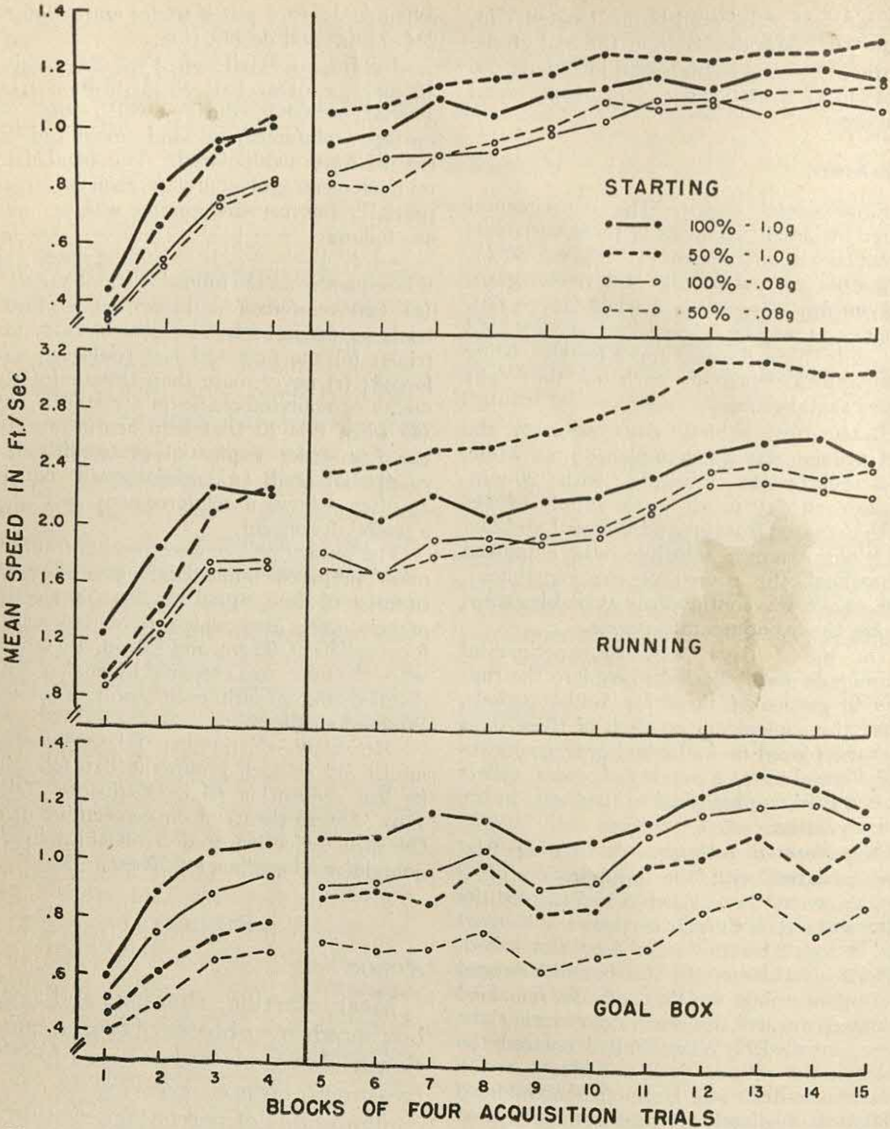


FIG. 1. Mean acquisition speeds for the treatment groups differing in magnitude of reward and percentage reinforcement. (The vertical line marks the point at which training is terminated for half the *S*s in each group.)

reinforcement produced faster speeds than did partial reinforcement.

In order to evaluate the reliability of these observations, as well as the equivalence of the 16- and 60-trial acquisition groups, a separate analysis of variance appropriate for a factorial design was computed for the means over Trials 1 to 16 for each of the three response measures. Since half of the *S*s in each of the groups were male and half female, sex was also included in these and all subsequent analyses as a "control" variable.

The results of these analyses showed that: (*a*) magnitude of reward was significant beyond the .005 level in the starting, running, and goalbox measures ($F = 26.67$, $F = 23.74$, and $F = 8.20$, respectively; $df = 1/112$); (*b*) percentage reinforcement was highly significant in the goalbox measure ($F = 48.35$, $df = 1/112$, $P < .001$), reached significance in the running measure ($F = 4.05$, $df = 1/112$, $P < .05$), but failed to approach significance in the starting measure ($F < 1$); (*c*) although the starting and running speed curves suggest an interaction between magnitude of reward and percentage reinforcement, this tendency was not statistically reliable in any of the response measures; and (*d*) there were no reliable differences between the experimental groups receiving only these 16 acquisition trials and those which subsequently received 44 additional trials.

With reference again to Fig. 1, it is apparent on all three response measures that each of the large reward groups maintained its advantage over the corresponding small reward group throughout the 60 acquisition trials. Likewise, in the case of the goalbox measure the 100% reinforcement groups maintained their advantage over the 50% groups. However, on both the starting and running measures the performance of each of the 50% groups may be seen to surpass that of the corresponding 100% group by the end of training.

In order to evaluate statistically the asymptotic performance levels attained under the various experimental treatments received by the 60-trial *S*s, a separate analysis of variance appropriate for a factorial design was computed for the means over the last 8 acquisition days on each of the response measures. The results of these analyses showed that: (*a*) the effect of magnitude of reward attained significance in both the starting and running measure ($F = 6.05$, $P < .025$, and $F = 9.53$, $P < .005$, respectively; $df = 1/56$) but is associated with a probability between .10 and .20 in the goalbox measure ($F = 2.67$, $df = 1/56$); (*b*) 50% reinforcement led to significantly higher performance levels than did 100% reinforcement on both the starting and running measure ($F = 4.03$, $P < .05$, and $F = 5.03$, $P < .025$, respectively; $df = 1/56$), whereas 50% reinforcement led to a significantly lower performance level than did 100% reinforcement on the goalbox measure ($F = 8.01$, $df = 1/56$, $P < .01$); and (*c*) in none of the three response measures was there a statistically reliable interaction of percentage reinforcement and magnitude of reward.

Extinction

Since the several experimental groups began extinction at markedly different levels, a comparison of the obtained extinction speeds appeared to offer only minimal information about the rate of extinction. A preferable index appeared to be extinction speed expressed as a proportion of the corresponding terminal acquisition speed. Consequently each response speed for each *S* over Extinction Trials 1–32 was expressed as a proportion of the mean speed *S* had attained on that response measure over the last four acquisition trials and Extinction Trial 0.

Figure 2 presents the mean starting, running, and goalbox proportions over blocks of four extinction trials for the eight experimental groups. Of major concern is the comparison of the resistance to extinction of the 50% and 100% groups following the various acquisition conditions. In all

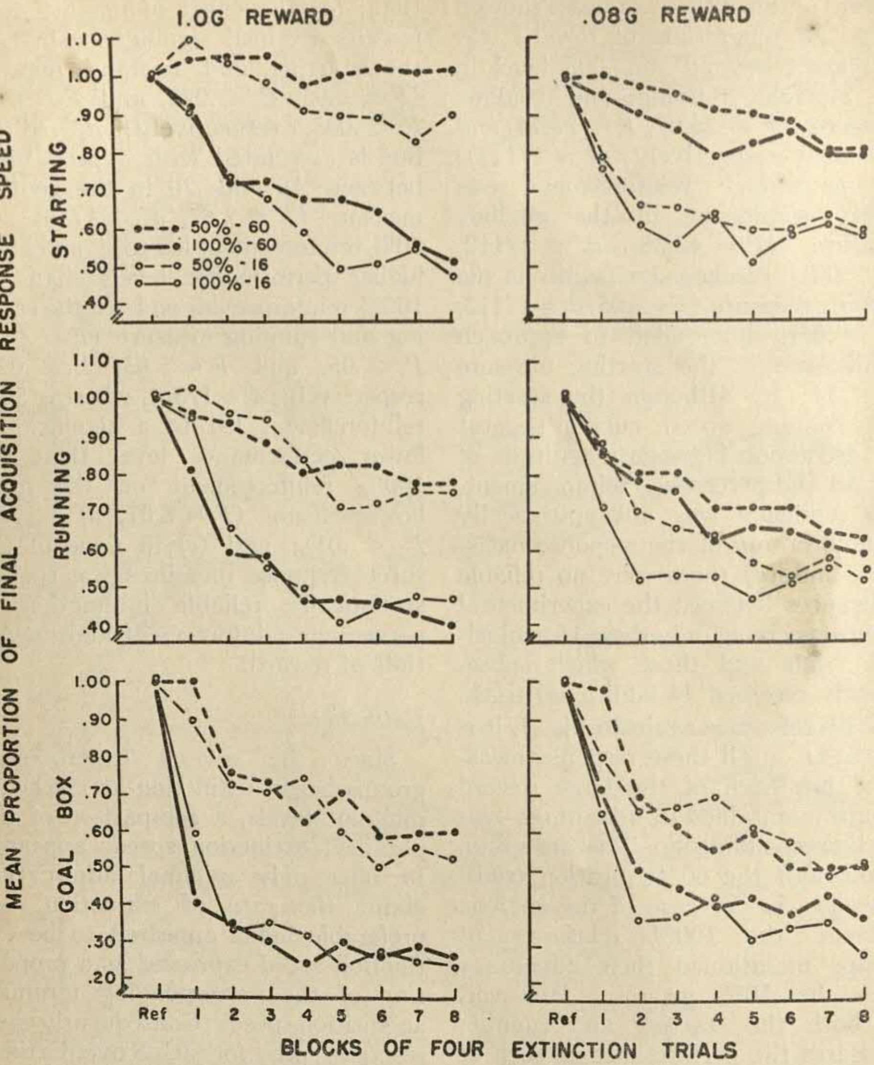


FIG. 2. Mean extinction proportions as a function of the percentage reinforced (50% or 100%) and the number (16 or 60) of prior acquisition trials plotted separately for the groups which had received 1.0-gm. or .08-gm. rewards during training. (The reference point includes the final four acquisition trials and Extinction Trial 0.)

three response measures it may be observed that with acquisition conditions of 1.0-gm. reward, 50% reinforcement led to a markedly greater resistance to extinction than did 100% reinforcement. However, when .08-gm. reward was employed, the superiority of the 50% groups, although still present in every comparison, was appreciably smaller. It should also be noted that in none of the comparisons did the difference in extinction performance between the 50% and 100% groups appear to vary with the number of prior acquisition trials.

A separate analysis of variance appropriate for a factorial design was computed for the mean proportions over Extinction Trials 1–32 for each of the response measures. The results of these analyses with respect to the percentage reinforcement variable showed that: (*a*) in none of the measures did the triple interaction of percentage reinforcement, magnitude of reward, and number of acquisition trials reach significance; (*b*) in none of the measures did the interaction of percentage reinforcement with number of acquisition trials reach significance ($F < 1$ in each case); but (*c*) in all three response measures the interaction of percentage reinforcement with magnitude of reward was statistically significant ($F = 15.70$, $P < .001$; $F = 14.97$, $P < .001$; and $F = 4.71$, $P < .05$ for the starting, running, and goalbox measures, respectively; $df = 1/112$).

Since the effects of percentage reinforcement were reliably different when combined with 1.0-gm. reward than with .08-gm. reward, separate F tests were computed of the differences between the 50% and 100% treatment groups at the two levels of reward magnitude. The results of these comparisons for the starting and running measures indicated that the 50% groups were significantly more resistant to extinction than the 100% groups when combined with 1.0-gm. reward ($F = 43.54$, and $F = 47.33$, respectively; $df = 1/112$; $P < .001$ in each case) but not when combined with .08-gm. reward ($F < 1$ and $F = 1.97$, respectively; $df = 1/112$). In the goalbox measure, however, percentage reinforcement was statistically reliable when combined with either 1.0-gm. reward ($F = 59.65$, $df = 1/112$, $P < .001$) or with .08-gm. reward ($F = 21.88$, $df = 1/112$, $P < .001$).

Also of interest are the separate effects of magnitude of reward and number of acquisition trials on resistance to extinction. As noted previously, magnitude of reward interacts with percentage reinforcement in determining resistance to extinction. Consequent to this finding, separate F tests were also computed comparing the 1.0-gm. and .08-gm. treatment groups at the two percentages of reinforcement on each of the three response measures. When combined with a 50% reinforcement schedule 1.0-gm. reward, as compared to .08-gm. reward, led to a greater resistance to extinction which attained statistical significance in the case of the starting measure ($F = 17.85$, $df = 1/112$, $P < .001$) and the running measure ($F = 10.69$, $df = 1/112$, $P < .005$), but not the goalbox measure ($F < 1$). When 100% reinforcement was employed, however, 1.0-gm. reward led to *less* resistance to extinction than did .08-gm. reward. The latter finding was significant in the case of the goalbox measure ($F = 5.71$, $df = 1/112$, $P < .025$) and the running measure ($F = 4.83$, $df = 1/112$, $P < .025$) but not in the starting measure ($F = 1.87$, $df = 1/112$, $.10 < P < .20$).

Statistical analyses revealed that an increase in the number of acquisition trials led to a reliable increase in the extinction proportions only in the case of those treatment groups which received .08-gm. reward, and then only on the starting and running measures ($F = 22.41$, $P < .001$, and $F = 5.72$, $P < .025$, respectively; $df = 1/112$) there being no reliable effect of number of acquisition trials upon the goalbox measure. Although number of acquisition trials seldom affected the extinction proportions it would still be expected to affect extinction *speeds*, since *S*s began extinction at a higher level following 60 than following 16 acquisition trials. When analyses of variance were performed on the untransformed extinction speeds significantly superior extinction performance of those groups receiving 60 acquisition trials, as compared to those receiving only 16 acquisition trials, was obtained in all comparisons.

DISCUSSION

Beier (1958) and Logan (1960) have assumed that under conditions of inconsistent reinforcement, $r_g - s_g$ is raised or lowered on each trial toward the value appropriate for the reward received on that trial. Several other investigators (e.g., Amsel, 1958; Spence, 1960, Ch. 6) have asserted, in addition to their individual assumptions concerning the course of $r_g - s_g$, that after some minimal number of reinforcements, nonreinforcement will elicit a primary motivational response termed *frustration*, which is directly related in intensity to the strength of $r_g - s_g$. An anticipatory or conditioned form (r_f) of this emotional response is further assumed to generalize to stimuli which antedate the goal, and is expected not only to lead to an increase in drive level, but also, through its characteristic response produced stimuli (s_f), to tend to elicit unlearned or previously learned responses which are antagonistic to the approach response. It is assumed however, that during random partial reinforcement training that these cues may become conditioned to, and tend to elicit the locomotor response itself.

While these positions when taken together lack in many instances the degree of development necessary for unequivocal deductions, they do allow for a consistent interpretation of a sizeable number of partial reinforcement phenomena. For example, during the early, prefrustration acquisition trials, partial reinforcement should lead to a lower $r_g - s_g$ than should continuous reinforcement and hence, as presently obtained, a lower level of responding on all response measures. If one makes the special assumption with Beier (1958) that $r_g - s_g$ under inconsistent reward, approaches the mean of the $r_g - s_g$s appropriate to the several rewards received, then $r_g - s_g$ under random 50% reinforcement approaches 50% of the level attained under 100% reinforcement and hence a greater partial reinforcement decrement should be obtained with large than with small reward. The present results are in this direction but do not attain significance. The fact that a similar nonsignificant tendency was observed by Hulse (1957) in the only previous attempt at such a comparison, however, suggests that it may be a reproducible effect.

The latter, postfrustration performance is less determinant, depending upon the relative degree of $r_g - s_g$ superiority of the continuous groups as compared to the $r_f - s_f$ induced drive level superiority of the partial groups, as well as upon the degree of conditioning of s_f to the approach response under partial reinforcement. Since random partial reinforcement leads commonly to greater asymptotic starting and running speeds than does continuous reinforcement, it must be argued not only that the approach response has become effectively conditioned to the early alley cues plus s_f, but also that the drive level advantage of the partial groups is greater than the $r_g - s_g$ advantage of the continuous groups.[3]

---

[3] That the partial superiority is not found on those response measures close to the goal may be attributed to less effective conditioning of the approach response to the goalbox cues plus s_f, and hence the persistance to some degree of competing responses elicited by s_f. This argument appears quite reasonable in the context of the prior assumptions if it is noted that $r_f - s_f$ may be expected to become conditioned to the approach response when the s_f cues are introduced initially at a weak value, and hence with a negligible tendency to elicit competing responses, and are then increased gradually at the same time that the approach response is being strengthened. This condition obtains much more clearly in the early portions of the alley than in the goal region where

Since $r_f - s_f$ is presumed to be a function of $r_g - s_g$, increasing the magnitude of reward should increase the partial reinforcement drive level superiority. The present finding of no significant asymptotic interaction between reinforcement schedule and reward magnitude might be interpreted as casting doubt on the frustration interpretation, or as lending additional plausability to the notion that the continuous reinforcement $r_g - s_g$ advantage is also increased with increased reward. However, that there was observed a sizeable and consistent tendency for the starting and running partial reinforcement asymptotic superiority to be greater with large than with small rewards appears to justify further study.

It follows most clearly from the theoretical formulation, that partially reinforced *S*s that have been trained to approach in the presence of frustration stimuli will be more resistant to extinction than will continuously reinforced *S*s that have competing responses elicited to a greater degree by the anticipatory frustration which develops during extinction. Furthermore, the present finding, as well as that of Hulse (1958), of a greater partial reinforcement extinction superiority following large than following small rewards is not unexpected. Nor is the observation in each of these studies, that this increased effect is due, in part, to *faster* extinction following continuous reinforcement with large as compared to small rewards. Since the magnitude of $r_g - s_g$ would be greater at the start of extinction with large than with small rewards, it is consistent that larger rewards should lead to a slower rate of extinction for partial *S*s due to increased motivation, but should lead to a faster rate of extinction for continuous *S*s owing to the greater vigor of the competing responses elicited by the more intense s_f.

On both theoretical (Amsel, 1958) and empirical (Amsel, 1958; Koehler, 1956; Weinstock, 1957) grounds it was expected that a greater partial reinforcement extinction superiority would be observed following 60 than following 16 acquisition trials. In order for the present failure to observe such a tendency to be considered consistent with frustration theory it must be proposed that s_f had already become effectively conditioned to running in the partially reinforced groups by the end of 16 trials. The present results would more clearly appear to place restrictions on the assumptions that may be made concerning the nature of the contributions of number of acquisition trials and magnitude of reward to the partial reinforcement extinction effect. While it may be assumed that an increase in the number of acquisition trials under certain conditions may lead to an increase in the magnitude of the partial effect, it must also be assumed that the function expressing this relationship has a limiting value determined by the magnitude of reward employed.

While it is not unlikely that both frustration and incentive motivation are important determiners of partial reinforcement phenomena, the present formulations are necessarily speculative. Further clarification awaits less equivocal evidence of the relative values of these factors in the various portions of training and at different segments of the response chain.

SUMMARY

The acquisition and extinction of locomotor behavior in various segments of a straight alley were investigated as a function of percentage reinforcement (50% or 100%), magnitude of reward (.08 gm. or 1.0 gm.), and number of acquisition trials (16 or 60). Sixteen naive hooded rats were assigned to each of the eight groups in a 2 × 2 × 2 experimental design. A 24-hr. intertrial interval was observed during acquisition and during the 33 extinction trials which followed.

The relative acquisition performance levels of partially and continuously reinforced *S*s were found to depend upon the portion of training and the segment of the response

---

due to the proximity of the primary frustration event, generalized $r_f - s_f$ may be expected to occur earlier and to follow a course of greater intensity, and thus produce responses which compete more effectively with the approach response.

chain that was observed. When the response measure was taken over the early segments (first and second 6 in.) of the 5-ft. alley, there was observed an initial superiority of the 100% *S*s but an asymptotic superiority of the 50% *S*s. When the response measure was taken over the final 6 in. preceding the goal cup the superiority of the 100% *S*s persisted over the entire course of acquisition. Large as compared to small rewards, in addition to producing higher overall response speeds, also tended to produce a greater early trials 50% decrement and a greater asymptotic 50% superiority, although no statistical support for the latter two tendencies was obtained.

Of primary interest during extinction were the effects of size of reward and number of acquisition trials on the magnitude of the commonly obtained superiority of partially reinforced as compared to continuously reinforced *S*s. The extinction results showed that whereas the "partial reinforcement effect" did not vary with number of prior acquisition trials, it was markedly greater following 1.0-gm. than following .08-gm. rewards. The increased partial reinforcement effect with larger rewards reflected not only greater resistance to extinction of partially reinforced *S*s but also less resistance to extinction of continuously reinforced *S*s with large as compared to small rewards.

The acquisition and extinction results were considered in relation to frustration and incentive motivation theory.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, 55, 102–119.

BEIER, E. M. Effects of trial-to-trial variation in magnitude of reward upon an instrumental running response. Unpublished doctoral dissertation, Yale University, 1958.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, 57, 57–63.

HAGGARD, D. F. Acquisition and extinction of a simple running response as a function of partial and continuous schedules of reinforcement. Unpublished doctoral dissertation, State University of Iowa, 1956.

HULSE, S. H. Schedule of reinforcement, amount of reinforcement, and duration of goal confinement as variables in conditioning and extinction. Unpublished doctoral dissertation, Brown University, 1957.

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, 56, 48–57.

KOEHLER, J., JR. Acquisition and extinction of a running response as a function of the percentage of reinforcement and the number of acquisition trials. Unpublished doctoral dissertation, Tulane University, 1956.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

SPENCE, K. W. *Behavior theory and learning: Selected papers.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WEINSTOCK, M. B. A factorial study of some variables affecting resistance to extinction under partial reinforcement with spaced trials. Unpublished doctoral dissertation, Indiana University, 1957.

(Received June 11, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 243–252

STIMULUS ATTRIBUTES AND DRIVE IN PAIRED–ASSOCIATE LEARNING[1]

HERBERT LEVITT[2] AND ALBERT E. GOSS

University of Massachusetts

The paradigm for acquired learning with a single paired-associate unit is homologous with the paradigm for classical conditioning (Goss, Morgan, & Golen, 1959; Spence, 1958). Because the number of units can be varied there are many combinations and patterns of relationships among stimulus and response members which evolve and which make paired-associate learning a very useful task for investigations of interactions of task with other variables. The present study was concerned with the single and joint effects on paired-associate learning of four task variables, the similarity and association values of stimulus and of response members, and of two presumed drive variables, manifest anxiety and failure stress.

Task Variables

Similarity.—Similarity of the stimuli of paired-associate units may be in a vertical fashion, within the stimulus or response members, or laterally between stimulus and response members. Such similarity has been defined in terms of common elements (Underwood, 1953a) such as the number of similar letters among the stimuli, and in terms of common responses (Gibson, 1942) and meaning (Underwood, 1953b). Increasing the numbers of common elements among stimulus members slows learning (Feldman & Underwood, 1957; Underwood, 1953a). While such increases among response members increase overt errors, no significant effects on acquisition speed has been reported (Underwood, 1953a).

Similarities of stimulus members defined by common responses is directly related to learning speed (e.g. Gibson, 1942). However, for similarity defined by common meanings, stimulus members of intermediate similarity may be learned more rapidly than those of either high or low similarity (Underwood, 1953b). Similarity of response members, in terms of common meaning, is directly related to overt errors, but may (Feldman & Underwood, 1957) or may not (Underwood, 1953b) be related inversely to learning speed.

Association values.—In general, association values of stimulus and of response members are directly related to learning speed (Jantz & Underwood, 1958; Lifton, 1956; Mandler & Huttenlocher, 1956). Cieutat, Stockwell, and Noble's (1958) results suggest that these relationships hold across all levels of association values of stimulus and of response members, while Mandler and Campbell's (1957) finding suggests that direct relationships occur at only some levels. On the whole, association values of response members seem to have greater effects than association values of stimulus members (Cieutat et al., 1958; Hunt, 1959; Noble & McNeely, 1957; Sheffield, 1946).

Sheffield's (1946) is the most systematic attempt to explain these largely direct relationships between learning speed and association values of stimulus and of response members. Stimulus members of high association values are presumed to elicit mediating perceptual responses which are more invariant, distinctive, and compatible than the perceptual responses aroused by stimulus members of lower association values. Furthermore, stimuli produced by the perceptual responses might be more important components of stimulus compounds to which responses are to be conditioned than the external stimuli which presumably elicit both perceptual responses and the responses to be learned.

---

[1] A more extensive account of the first of the two experiments reported here can be found in Levitt (1956). Parts of these data were presented at the 1957 meetings of the EPA in New York City. Experiment II was supported by a grant from the Research Council of the University of Massachusetts. Herbert Gewirtz, Harvey Lifton, and Ira Mintz contributed to various aspects of the two experiments.

[2] Now at Department of Psychiatry, University of Wisconsin Medical School, and Wisconsin Psychiatric Institute, University Hospitals, Madison, Wisconsin.

These postulated aspects of association values of stimulus and of response members of paired-associate units are independent of similarity defined by common elements. Similarity and association values of stimulus members might, therefore, combine to influence paired-associate learning. Also, the manner of combination of these variables could well be different with response members. In order to obtain this information, in Exp. I similarity and association values of stimulus members were varied in a 2 × 2 orthogonal design, and in Exp. II similarity and association values of response members were varied in a 2 × 2 orthogonal design.

Drive Variables

Anxiety and stress.—Variations in scores on the Taylor Manifest Anxiety scale have been used to specify levels of drive during the acquisition of stimulus-response relationships of both simple and more complex tasks (Spence, 1958). Verbal statements of failure—failure stress—have been used to induce presumed drive levels. In Exp. I high and low levels of drive defined by high and low scores on the Taylor Manifest Anxiety scale were combined with stress and nonstress. In Exp. II, the levels of drive were defined only by high and low *MA* scores. While the manner of combination of these two specifications of drive was of some interest, the more important concern was the modes of combination of these drive variables with task variables of paired-associate learning.

Task variables and drive variables.—The two most important contemporary formulations of the possible modes of combination of task variables of paired-associate learning and of drive variables are what are here labeled the drive-dominant initial response analysis and the interference analysis. The drive-dominant initial response analysis as developed by Spence predicts that when correct responses to stimulus members of paired-associate units are stronger initially than incorrect responses, more rapid initial increases in performance should occur under high than under low drive. When incorrect responses are stronger initially than correct responses, better initial performance should be obtained under low than under high drive. The interference analysis (Child, 1954; Farber, 1955; Spence, 1958; Taylor, 1958) assumes that increases in drive increase the numbers and strengths of extralist and perhaps intralist responses which compete with correct responses to stimulus members. Accordingly, regardless of the initial relative strengths of correct and incorrect responses, this analysis predicts an inverse relationship between learning speed and drive.

Spence and his associates (Besch, 1959; Spence, 1958; Spence, Farber, & McFann, 1956; Spence, Taylor, & Ketchel, 1956) have constructed lists of paired-associates in which there is initial superiority of correct to incorrect responses or of incorrect to correct responses. While these investigators have obtained better results with the former than with the latter lists, neither Buchwald (1959) nor Kamin and Fedorchak (1957) obtained the predicted reversal.

The more common procedure for the construction of lists of paired-associates is to select stimuli for which there are no or minimal pre-established direct associations between stimulus and response members. Spence has already noted one difference between the two procedures;

when lists have been constructed with no pre-established associations between stimulus and response members, one or more trials must occur before correct or intralist incorrect responses to a stimulus member gains initial dominance. Because of this and other differences, findings with lists constructed by Spence's procedure would be expected to differ from those with lists constructed by the conventional procedure. Thus, what is required is direct determination of the mode of combination of drive variables with characteristics of stimuli of paired-associate lists constructed by the conventional procedure. In Exp. I of this study, the 2 × 2 orthogonal variations in similarity and association values of stimulus members were combined with the 2 × 2 orthogonal variations in manifest anxiety and stress. In Exp. II, the 2 × 2 orthogonal variations of similarity and association values of response members were administered to high-anxiety and low-anxiety *S*s. (Because too few *S*s were available at the time this experiment was undertaken, stress was not used.)

With stimulus members and with response members of low similarity and high association values correct responses might soon become dominant over incorrect responses. With stimulus members and with response members of high similarity and low association values, after a few trials, incorrect responses might be temporarily stronger than correct responses. If so, the implication of the drive-dominant initial response analysis is of better performance under high than under low drive with stimulus and with response members of low similarity and high association values. Conversely, with stimulus and with response members of high similarity and low association values poorer initial performance is predicted under high than under low drive. The interference analysis predicts a direct relationship between retardation of performance and drive early and probably late in learning, regardless of similarity and association values of stimulus or of response members.

EXPERIMENT I

Method

Paired-associates.—The stimulus members of the four lists of eight pairs of three-letter nonsense syllables were used. The stimulus members were one of four combinations of high or low similarity with high or low association values. Both sets of stimulus members of high similarity were made up of six consonants and three vowels. The Glaze association values of the set with high association values were 93% and 100% ($M = 97.4\%$) and those of the set with low association values were 0% and 7% ($M = 2.6\%$). The syllables of both low-similarity sets were made up of 16 consonants and four vowels. One set had mean Glaze values of 98.2% and the other had mean Glaze values of 2.6%.

The four sets of stimulus members were each paired with the same set of response members whose 15 different letters made them of intermediate similarity. Their 47% and 53% association values ($M = 49.2\%$) were also intermediate.

A modified Saltz and Myers' (1955) group presentation procedure was employed. Each of the eight paired-associates of each of the lists was first reproduced in pica capitals on 4¼ × 2¾ in. sheets of paper. Stimulus members of pairs were on the left and response members were on the right. In order to minimize serial effects, the eight sheets for the paired-associates of each list were assembled in four different random orders. Each of the four orders for each list was then stapled as a booklet in one of the four quadrants of a rectangular (7 × 11 in.) piece of posterboard. A vertical column of only stimulus members of the paired-associates was typed in pica capitals on each of the 4 × 5½ in. pages of a booklet on the reverse side. Any one of the four orders in which stimulus and response members appeared on a given trial was always followed by stimulus members alone in the same order.

Procedure for learning.—The four lists were administered simultaneously to groups of about 10 *S*s. The part of the instructions which was common for all groups and for all lists explained the design and use of the booklets. Each trial began by telling *S*s the booklet for that trial. They then waited for *E* to say "begin," whereupon the cover sheet of the appropriate booklet was opened. Each pair of syllables was examined by *S*s with the intention of later being able to write each of the syllables on the right opposite the syllable on the left with which it had been paired. Four seconds were allowed for the examination of each pair before the signal to turn to the next pair. After viewing the last pair, *S*s turned their posterboards over and waited for the signal to lift the cover sheet of the booklet on that side and to begin writing the correct responses to stimulus members alone. They had 32 sec. to write those responses after which they tore out that page, wrote their names on it, and passed it to a second *E*. The posterboard was then turned back to the first side. During the next trial the second *E* pretended to be scoring the answer sheet.

Twenty learning trials were administered to all groups with each of the four orders of the pairs of each list occurring five times in an otherwise random sequence.

Procedure for stress.—Stress was defined by two conditions: a statement to *S*s before beginning that the paired-associate task and academic performance were related, and subsequent statements of failure after Trials 6 and 10. After Trial 6 the second *E* left the table where he was scoring, came in front of the group, and announced, "As you have probably noticed, we have been scoring your test forms. Aren't you people trying? Maybe you're not cooperating. In comparison to similar groups who have taken this test previously, you people are doing significantly poorer. Now, let's try to put in a little more effort." A similar statement was made after Trial 10. The person who served as the second *E* was of mature appearance, considerably above average in height and size, and had a relatively deep, masculine voice.

The same *E* also interrupted nonstress *S*s after Trial 6 and 10 but with innocuous remarks of about the same lengths as the failure statements. These remarks controlled for the possibility that failure statements merely distracted. Upon completion of the learning task, *S*s of the stress condition were told the purpose of their experiences and asked not to describe them to anyone else.

Subjects.—The Taylor Manifest Anxiety scale (1953) was administered to students in introductory psychology during the first and second semesters of the 1955–56 academic year. Eighty high-anxiety *S*s and 80

TABLE 1

MEANS AND *SD*s OF CORRECT RESPONSES FOR TRIALS 1–20 IN EXPERIMENTS I AND II

Treatment Combinations				Experiment I (Stimulus Members)		Experiment II (Response Members)	
Similarity	Association Values	Anxiety	Stress	Mean	*SD*	Mean	*SD*
High	High	High	Yes	92.8	42.5	111.8	23.6
High	High	High	No	101.9	26.0		
High	High	Low	Yes	91.5	39.2	113.9	25.0
High	High	Low	No	86.1	26.8		
High	Low	High	Yes	66.0	36.5	75.9	28.4
High	Low	High	No	89.0	25.6		
High	Low	Low	Yes	61.9	24.7	76.9	33.5
High	Low	Low	No	73.8	25.6		
Low	High	High	Yes	120.6	13.1	125.0	27.6
Low	High	High	No	103.0	20.0		
Low	High	Low	Yes	115.7	20.0	125.1	18.3
Low	High	Low	No	124.2	24.5		
Low	Low	High	Yes	110.3	27.7	99.5	29.2
Low	Low	High	No	91.1	25.6		
Low	Low	Low	Yes	86.1	29.4	94.8	36.5
Low	Low	Low	No	99.6	21.9		

Note.—$N = 10$ in each combination.

TABLE 2

ANALYSIS OF VARIANCE OF MEANS FOR SUCCESSIVE FIVE-TRIAL BLOCKS IN EXPERIMENTS I AND II

Source	Experiment I (Stimulus Members)			Experiment II (Response Members)		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *S*s	159			79		
Similarity (Sim)	1	5,499.02	25.67**	1	1,357.13	6.11*
Association values (AV)	1	3,900.62	18.21**	1	5,176.15	23.32**
Anxiety (A)	1	200.26	—	1	0.70	—
Stress (St)	1	88.51	—			
Sim × AV	1	4.23	—	1	91.38	—
Sim × A	1	213.91	—	1	18.53	—
Sim × St	1	445.56	2.08			
AV × A	1	182.76	—	1	10.88	—
AV × St	1	187.06	—			
A × St	1	172.22	—			
Sim × AV × A	1	138.75	—	1	4.27	—
Sim × AV × St	1	120.75	—			
Sim × A × St	1	1,113.02	5.20*			
AV × A × St	1	15.62	—			
Sim × AV × A × St	1	1.57	—			
Error (b)	144	214.23		72	221.94	
Within *S*s	480			240		
Blocks of Trials (T)	3	15,088.05	858.25**	3	6,275.45	—
T × Sim	3	218.34	12.42**	3	6.28	—
T × AV	3	18.84	1.07	3	79.86	4.72**
T × A	3	14.15	—	3	4.93	—
T × St	3	42.95	2.44			
T × Sim × AV	3	20.17	1.14	3	52.24	3.09*
T × Sim × A	3	2.84	—	3	9.94	—
T × Sim × St	3	70.51	4.01**			
T × AV × A	3	8.42	—	3	45.13	2.67*
T × AV × St	3	3.04	—			
T × A × St	3	24.68	1.40			
T × Sim × AV × A	3	64.72	3.68*	3	6.22	—
T × Sim × AV × St	3	5.91	—			
T × Sim × A × St	3	51.04	2.90*			
T × AV × A × St	3	13.57	—			
T × Sim × AV × A × St	3	5.46	—			
Error (w)	432	17.58		216	16.91	

* Significant at .05 level.
** Significant at .01 level.

low-anxiety *S*s were drawn from among those *S*s whose manifest anxiety scores were, respectively, in the upper fifth (scores ≥ 20) and lower fifth (scores ≤ 10) of the distribution of all scores. Ten *S*s from each of these two groups of 80 *S*s were then assigned randomly to each of the eight combinations of high or low similarity, high or low association values, and stress or nonstress.

Results

Table 1 contains means and *SD*s of numbers of correct responses for all 20 trials for the 16 combinations of stimulus characteristics and drive conditions. Differences among means were assessed by the analysis of variance summarized in Table 2. This analysis includes a breakdown of the means for the 20 trials into means for successive blocks of 5 trials which do not appear in Table 1 but are the Trials (T) variable of the within-*S*s components in Table 2.

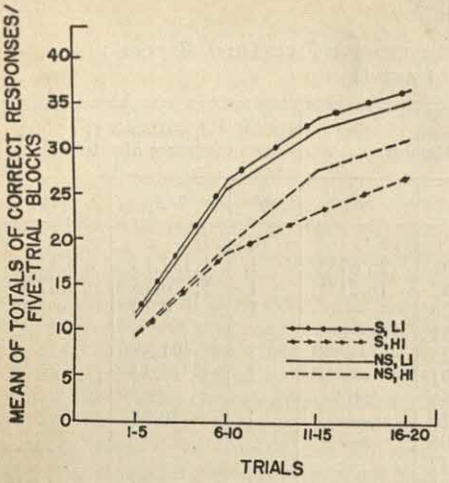


FIG. 1. Means of correct responses in successive blocks of five trials for combinations of stress (S) and nonstress (NS) with lists of high (HI) and low (LI) intralist similarity of stimulus members.

The main effects for similarity (Sim) and association values (AV) were significant beyond the .05 level. Summing over the other variables, the mean of 106.3 correct responses for 20 trials with stimulus members of low similarity was about 30% greater than the mean of 82.9 for stimulus members of high similarity. Stimulus members with high association values (M = 104.5) had nearly as great an advantage over those with low association values (M = 84.7). The *F*s for stress, anxiety, and their interactions were not significant.

Stress and similarity entered into three interactions significant at the .01 and .05 levels. The curves in Fig. 1 for the similarity, stress, and trials interaction illustrate the pattern of these interactions. The four curves diverge most during the last 10 trials. During these trials, the curve for low similarity lists rose less rapidly under the nonstress condition than that under stress. Both curves begin and remain above those for the high similarity lists on which stress produced some retardation. In other interactions involving one or more of the other variables with stress and similarity the same pattern held. Stress proved to be facilitative with lists in which stimulus members were of low similarity and inhibitory with lists in which stimulus members were of high similarity.

EXPERIMENT II

Method

Paired associates.—Stimulus and response members of the pairs of the four lists of Exp. I were interchanged to obtain the lists of this experiment. Thus, stimulus members were of intermediate similarity and association values and response members were combinations of high or low similarity and high or low association values. In all other respects, materials for the administration of these lists paralleled those of Exp. I.

Procedure for learning.—Except for the omission of the stress-nonstress condition the procedure for learning the paired-associates of Exp. I was followed.

Subjects.—Eighty new *S*s were drawn from among students in introductory psychology during the second semester of 1956–57. Half had manifest anxiety scores which placed them in the upper fifth of the distribution of those scores; the other half had scores which placed them in the lower fifth. The cutoff scores were those of Exp. I.

Ten of the *S*s with high anxiety and 10 of those with low anxiety were assigned randomly to each of the four combinations of similarity and association values of response members.

Results

Experiment II.—Means and *SD*s of numbers of correct responses for all 20 trials for the eight combinations of characteristics of response members and manifest anxiety appear in Table 1. With the exception of the stress-nonstress variable the analysis of variance for differences among these means was the same as that for Exp. I (Table 2).

The *F*s for the main effects of similarity (Sim) and association value

(AV) were, as in Exp. I, significant beyond the .05 level. The mean for correct responses of response members of low similarity ($M = 111.1$) was greater than that for high similarity ($M = 94.6$). Those for high ($M = 119.0$) and low ($M = 86.8$) association values also differed significantly. The main effect for manifest anxiety was not significant, nor did this variable enter into any significant interactions whose patterns were consistent.

Experiments I and II.—The effects of manipulating the stimulus and response sides on the learning of paired-associates were compared by analysis of variance (Table 3). (Manifest anxiety was not included as a factor because it had no apparent effects in Exp. I and, at most, limited effects in Exp. II.)

Significantly more correct responses were made during the 20 trials with variations in response members ($M = 102.9$) than with variations in stimulus members ($M = 94.8$). As expected from the separate analyses, low similarity of stimulus and response members together ($M = 107.9$) led to faster learning than did high similarity of both members together ($M = 86.8$), as was the case for high association values of stimulus and response members together ($M = 109.3$) in relation to low association values together ($M = 85.4$). The difference of 23.4 between means for stimulus members of low ($M = 106.3$) and high ($M = 82.9$) similarity was greater than the difference of 16.5 between means for response members of low ($M = 111.1$) and high ($M = 94.6$) similarity. For stimulus members of high ($M = 104.5$) and low ($M = 84.7$) association values, the difference of 19.8 was less than the difference of 32.2 for response members of high ($M = 119.0$) and low ($M = 86.8$) association values. The suggestion is that similarity had a greater effect on stimulus than on response members; and the association values had a greater effect on response than on stimulus members. Yet, these differences were not sufficient to produce a significant interaction of either similarity or association values with the member varied. Nor was the interaction of both similarity and association values with the member varied significant.

TABLE 3

ANALYSIS OF VARIANCE OF MEANS FOR SUCCESSIVE FIVE-TRIAL BLOCKS FOR SIMILARITY AND ASSOCIATION VALUES OF STIMULUS MEMBERS AND RESPONSE MEMBERS

Source	*df*	*MS*	*F*
Between *S*s	239		
Response or stimulus member (M)	1	910.25	4.24*
Similarity (Sim)	1	6,693.98	31.22**
Association Value (AV)	1	8,562.17	39.93**
M × Sim	1	162.17	—
M × AV	1	514.60	2.40
Sim × AV	1	51.81	—
M × Sim × AV	1	43.81	—
Error (b)	232	214.41	
Within *S*s	720		
Blocks of trials (T)	3	21,310.25	1,199.23**
T × M	3	53.25	3.00*
T × Sim	3	164.82	9.28**
T × AV	3	75.36	4.24**
T × M × Sim	3	59.81	3.36*
T × M × AV	3	23.34	1.31
T × Sim × AV	3	57.59	3.24*
T × M × Sim × AV	3	14.82	—
Error (w)	696	17.77	

* Significant at .05 level.
** Significant at .01 level.

DISCUSSION

Similarity and association values.—The slower learning with stimulus members of high similarity than with those of low similarity was consistent with previous findings with nonsense syllables (Underwood, 1953a). However, that this relationship held for similarity of response members and there was no significant difference in the effects of similarity of stimulus and response members were not in accord with Underwood's findings. Underwood used lists in which similarity of both stimulus and response members were varied; in the present study similarity of either stimulus members or of response members but not of both members was varied. Also, all of his nonsense syllables were of intermediate association values.

The combinations of association values of stimulus and response members of the lists of this study are most comparable to those lists in Mandler and Campbell's (1957) experiment, and in Cieutat, Stockwell, and Noble's (1958) Exp. II. These lists had stimulus members of different association values and response members of intermediate association values or stimulus members of intermediate association values and response members of different association values. For these lists, Mandler and Campbell found that learning speed was related to association values of both stimulus and response members by an inverted U function in their Exp. I and by a U function in their Exp. II. In contrast, the results of this study and those of Cieutat, Stockwell, and Noble suggest a direct monotonic relationship between learning speed and association values of both stimulus and response members. On the whole, the results of other studies suggest that the relationships for association values of stimulus members, response members or both are direct and monotonic (e.g., Feldman & Underwood, 1957; Hunt, 1959; L'Abate, 1959; Noble & McNeeley, 1957; Sheffield, 1946).

The results of this study suggested that association values of stimulus and of response members had essentially the same effects on learning speed. In contrast, Cieutat, Stockwell, and Noble (1958) found that association values of response members had significantly greater effects than association values of stimulus members. Their results are more typical (e.g., Hunt, 1959; Noble & McNeely, 1957; Sheffield, 1946). It is probable that one or more of several differences in the conditions of the present study and the conditions of these other experiments accounts for this difference in results. In the latter experiments, responses were pronounced rather than spelled, or both members of the pairs were of high or low association values rather than one member of intermediate association values, or stimulus or response members of varying association values appeared within the same lists rather than in different lists.

Task variables and drive.—The results of Exp. I suggested with stimulus members of low similarity (easier) stress was facilitative and with stimulus members of high similarity (harder) stress was inhibitory. Thus, whether stress is facilitative or inhibitory depends on the characteristics of the lists. This is a finding which is in accord with implications of the drive-dominant initial response analysis, but not with that of the interference analysis of the role of drive in paired-associate learning.

That stress interacted with similarity of stimulus members but not with their association values supports Sheffield's (1946) proposal of different bases for the effects of similarity and association values on acquisition. High similarity in terms of common physical elements may increase interstimulus generalization sufficiently to bring about initial dominance of intralist incorrect responses rather than of correct responses.

The relatively slower learning which characterizes lists with stimulus members of low rather than of high association values, regardless of their similarity, is probably not based on interstimulus generalization. Responses of repeating stimuli of low association values presumably have longer and more variable latencies and durations and more variable forms than such responses to stimuli with high association values. Such greater variability for both stimulus and response members in stimuli produced by responses of repeating the low than for those of high association values probably accounts for the direct relationships between learning speeds and association values of both stimulus and response members.

There was no indication that anxiety interacted with similarity either of stimulus or of response members. In contrast, Taylor and Chapman (1955) and Taylor (1958) reported more rapid learning of paired-associates of low intralist similarity and low association values by *S*s with high than with low anxiety. There is no simple explanation of the lack of agreement between these findings and those of the present study. Differences in the lists may be one factor. Also, the group procedure of administering the paired-associates used here may have been sufficiently similar to classroom test situations to raise the anxiety levels of all *S*s to about the same high levels, regardless of their manifest anxiety scores.

In Exp. II, the interaction of anxiety, association values, and trials was just significant at the .05 level. Other interactions involving these variables were not significant. Therefore, the results of this study seem best interpreted as consistent with L'Abate's (1959) finding of no interaction of anxiety and association values. Thus, this aspect of the results is not in accord with implications of either the drive-dominant initial response or interference analysis.

SUMMARY

In Exp. I, four lists of eight paired-associates representing 2 × 2 orthogonal variations in similarity (number of letters in common) and association values (Glaze) of nonsense-syllables stimulus members were learned under presumed drive conditions indicated by high or low Taylor Manifest Anxiety scores combined with failure stress or nonstress. The nonsense-syllable response members of these lists were of intermediate similarity and intermediate association values. In Exp. II, the stimulus and response members of the four lists were reversed so that the response members represented 2 × 2 orthogonal variations in similarity and association values. These lists were learned by *S*s with high or low anxiety scores.

For both experiments lists were administered for 20 learning trials by a group procedure. The failure stress of Exp. I was introduced after Trials 6 and 10.

Learning speeds varied directly with association values of both stimulus and response members, and inversely with similarity of both stimulus and response members. Stress seemingly facilitated acquisition of lists with stimulus members of low similarity relative to those with high similarity and retarded acquisition of lists with stimulus members of high similarity relative to those with low similarity. Stress had no effects on acquisition of stimulus members of high or low association values; anxiety had no effects on acquisition of response members of varying similarity or association values either alone or in combination. The form of interaction of stress with similarity was consistent with Spence's drive-dominant initial response analysis rather than with an interference analysis of the role of drive in complex tasks. However, this was the only finding which supported either analysis.

REFERENCES

BESCH, N. F. Paired-associates learning as a function of anxiety level and shock. *J. Pers.*, 1959, **27**, 116–124.

BUCHWALD, A. M. Manifest anxiety level, verbal response-strength, and paired-associates learning. *Amer. J. Psychol.*, 1959, **72**, 89–93.

CHILD, J. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m*, *m'*) in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

FARBER, I. E. Motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–325.

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **58**, 11–15.

GIBSON, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.

GOSS, A. E., MORGAN, C. H., & GOLIN, S. J. Paired-associates learning as a function of percentage of occurrence of response members (reinforcement). *J. exp. Psychol.*, 1959, **52**, 96–104.

HUNT, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, **57**, 262–267.

JANTZ, E. M., & UNDERWOOD, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, **56**, 174–179.

KAMIN, L. J., & FEDORCHAK, O. The Taylor scale, hunger, and verbal learning. *Canad. J. Psychol.*, 1957, **11**, 212–218.

L'ABATE, L. Manifest anxiety and the learning of syllables with different associative values. *Amer. J. Psychol.*, 1959, **72**, 107–110.

LEVITT, H. Paired-associates learning as a function of manifest anxiety, stress, intralist similarity, and stimulus association value. Unpublished master's thesis, University of Massachusetts, 1956.

LIFTON, H. An experimental test of the occurrence of intermodal transfer in human verbal learning. Unpublished master's thesis, University of Massachusetts, 1956.

MANDLER, G. H., & CAMPBELL, E. H. Effects of variations in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 269–273.

MANDLER, G. H., & HUTTENLOCHER, J. The relationship between associative frequency, associative ability, and paired-associate learning. *Amer. J. Psychol.*, 1956, **69**, 424–428.

NOBLE, C. E., & MCNEELY, D. A. The role of meaningfulness (*m*) in paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 16–23.

SALTZ, E., & MYERS, T. I. A method for group presentation of paired-associates learning materials. Paper read at Midwestern Psychological Association, Chicago, 1955.

SHEFFIELD, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctoral dissertation, Yale University 1946.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., FARBER, J. E., & MCFANN, H. A. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, J. A. The effects of anxiety level and psychological stress on verbal learning. *J. abnorm. soc. Psychol.*, 1958, **57**, 55–60.

TAYLOR, J. A., & CHAPMAN, J. P. Anxiety and the learning of paired-associates. *Amer. J. Psychol.*, 1955, **68**, 671.

UNDERWOOD, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intra-list similarity. *J. exp. Psychol.*, 1953, **45**, 133–142. (a)

UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intra-list similarity. *J. exp. Psychol.*, 1953, **45**, 143–149. (b)

(Received July 5, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 253–262

THE OBSERVING RESPONSE IN DISCRIMINATION LEARNING[1]

RICHARD C. ATKINSON
Stanford University

Recently, the importance of an orienting or observing response has been emphasized in the formulation of a general theory of discrimination behavior (Atkinson, 1958, 1960; Burke & Estes, 1957; Restle, 1959; Wyckoff, 1952). Unfortunately, for many experimental problems it is not clear how such a theory should be formalized. In particular, there are not enough experimental data available to permit a detailed specification of the postulates relating observing responses and such variables as stimulus dimensions, reinforcement schedules, and stimulus schedules. The purpose of this study is to gain information about this class of relations by modifying the typical discrimination task so that observing responses can be categorized and directly measured.

The experimental situation is considered as a sequence of discrete trials. Each trial is described in terms of the following classifications:

T_1, T_2: *Trial type.* Each trial is either a T_1 or a T_2. Trial type is selected by E and determines in part the stimulus event occurring on that trial.

O_1, O_2: *Observing responses.* At the start of each trial, S makes either an O_1 or O_2. The particular observing response made determines in part the stimulus event for that trial.

s_1, s_c, s_2: *Stimulus events.* Following the observing response, one and only one of these stimulus events (discriminative cues) occurs. On a T_1 trial, s_1 or s_c can occur; on a T_2 trial, s_2 or s_c can occur.

A_1, A_2: *Discrimination responses.* On each trial S makes either an A_1 or A_2 response to the presentation of the stimulus event.

E_1, E_2: *Reinforcing events.* The trial is terminated with the occurrence of one of these events. An E_1 indicates that A_1 was the correct response for that trial and E_2 indicates that A_2 was correct.

The sequence of events on a trial is as follows: (*a*) ready signal occurs and S makes either an O_1 or O_2; (*b*) following the observing response s_1, s_2, or s_c is presented; (*c*) to the onset of s_i, S makes either A_1 or A_2; (*d*) the trial is terminated with reinforcing event E_1 or E_2.

The trial type and reinforcing event are determined by E. The probability of an E_1 event on a T_1 trial is denoted by π_1, and the probability of an E_1 event on a T_2 trial is denoted by π_2. Consequently, the probability of an E_2 is $1 - \pi_1$ on a T_1 trial and $1 - \pi_2$ on a T_2 trial. The two types of trials are equiprobable in the present experiment.

The particular s_i event that is presented on any trial depends on the trial type and the observing response. If an O_1 is made, then (*a*) with probability α the s_1 event occurs on a T_1 trial and the s_2 event occurs on a T_2 trial, and (*b*) with probability $1 - \alpha$ the s_c event occurs, regardless of the trial type. If an O_2 is made, then (*a*) with probability α the s_c event occurs, regardless of the trial type, and (*b*) with probability $1 - \alpha$ the s_1 event occurs on a T_1 trial and the s_2 event occurs on a T_2 trial.

To clarify the experimental procedure, consider a case where $\alpha = 1$, $\pi_1 = 1$, and $\pi_2 = 0$. If S is to be correct on every trial, he must make an A_1 on a T_1 trial and an A_2 on a T_2 trial. However, S can gain information about the trial type only by making the appropriate observing response. That is, O_1 must be made in order to identify the trial type; the occurrence of O_2 always leads to the presentation of s_c. Hence, for perfect responding in this case, S must make the O_1 response with probability 1 and then make A_1 to s_1 or A_2 to s_2.

---

[1] This research was supported by the Office of Naval Research under Contract Nonr 233(58).

The aim of this study is to investigate the effect of various event schedules on observing behavior. In particular, we are interested in the values of π_1, π_2, and α as determiners of the probability of an O_1 response.

THEORY

The analysis of the data will be organized within the framework of a Markov chain model which is closely related to stimulus sampling theory as first formulated by Estes (1950) and Estes and Burke (1953). The mathematical techniques for the model considered in this paper have been presented in detail elsewhere (Atkinson, 1960; Suppes & Atkinson, 1960) and the reader is referred there for a rigorous development.

The basic assumption for observing responses is that if $O_i (i = 1, 2)$ occurs and leads to the selection of a stimulus which in turn elicits a correct discrimination response, then *S* will tend to repeat that observing response on the next trial. However, if O_i occurs and leads to the selection of a stimulus which elicits an incorrect discrimination response, then *S* will tend not to repeat that observing response on the next trial. Conceptually, this assumption is similar to that proposed by Wyckoff (1952) and Atkinson (1958).

It is next assumed that *S* can be described by an ordered four-tuple at the start of trial n where (*a*) the first member is 1 or 2 and indicates whether O_1 or O_2 will be made on trial n, (*b*) the second member is 1 or 2 and indicates whether s_1 is conditioned to A_1 or to A_2 (i.e., whether A_1 or A_2 will occur if s_1 is presented), (*c*) the third member is 1 or 2 and indicates whether s_c is conditioned to A_1 or to A_2, and (*d*) the fourth member is 1 or 2 and indicates whether s_2 is conditioned to A_1 or to A_2.

These four-tuples will be referred to as subject states and assigned identifying numbers as follows:

1. (1111)
2. (1112)
3. (1121)
4. (1122)
5. (1211)
6. (1212)
7. (1221)
8. (1222)
9. (2111)
10. (2112)
11. (2121)
12. (2122)
13. (2211)
14. (2212)
15. (2221)
16. (2222)

From trial to trial *S* may change states depending on the sequence of responses and reinforcements. The possible changes are specified by the following axioms:

Axiom 1: With probability θ' the $s_k (k = 1, 2, c)$ stimulus presented on trial n will become conditioned to the reinforced response; if it is already conditioned to that response it remains so. (For example, if s_k is presented and followed by E_j then with probability θ' it will become conditioned to A_j.)

Axiom 2: If $O_i (i = 1, 2)$ is made on trial n and followed by an s_k which elicits a correct discrimination response, then *S* will repeat the same observing response on the next trial. However, if O_i is made and followed by an incorrect discrimination response, then with probability θ'' *S* will make the other observing response on the next trial.

From these assumptions and the event schedules employed in this experiment, it can be shown that the sequence of random variables which take the subject states as values is an irreducible, aperiodic Markov chain. This means among other things that a transition matrix $[p_{ij}]$ may be derived from these assumptions where p_{ij} is the conditional probability of being in state j on trial $n + 1$ given state i on trial n. The learning process is completely characterized by these transition probabilities and the initial probability distribution on states.

To clarify the application of the axioms we derive one element of $[p_{ij}]$. Assume *S* is in State 1211 at the start of Trial n and T_1 E_1 is selected by *E* with probability $\frac{1}{2}\pi_1$. Then an O_1 occurs with probability 1 and an s_1 is presented with probability α; to the presentation of s_1 an A_2 is made. The *S*'s discrimination response was incorrect and therefore with probability θ'' the observing response changes from O_1 to O_2. Also, with independent probability θ' the conditioning of s_1 changes from A_2 to A_1. Multiplication of the conditional probabilities yields the probability of going from State 1211 to State 2111; i.e., $p_{5,9} = \frac{1}{2}\pi_1\alpha\theta'\theta''$.

In this paper, we shall be primarily interested in the asymptotic behavior of

S. Consequently, $p_{ij}^{(n)}$ is defined as the probability of being in state j on trial $n+1$, given that on Trial 1 S was in State i. Then the following limit exists and is independent of i,

$$u_j = \lim_{n \to \infty} p_{ij}^{(n)}.$$

The quantity u_j can be interpreted as the asymptotic probability of being in state j no matter what the initial distribution. Experimentally, we will be interested in evaluating the following theoretical predictions:

$$P_\infty(O_1) = u_1 + u_2 + u_3 + u_4 + u_5 + u_6 + u_7 + u_8 \quad [1]$$

$$P_\infty(A_1 | T_1) = u_1 + u_2 + u_9 + u_{10} + \alpha[u_3 + u_4 + u_{13} + u_{14}] + (1-\alpha)[u_5 + u_6 + u_{11} + u_{12}] \quad [2]$$

$$P_\infty(A_1 | T_2) = u_1 + u_5 + u_9 + u_{13} + \alpha[u_3 + u_7 + u_{10} + u_{14}] + (1-\alpha)[u_2 + u_6 + u_{11} + u_{15}] \quad [3]$$

$$P_\infty(O_1 \cap A_1) = u_1 + \alpha u_3 + (1-\alpha)u_6 + (\alpha/2)[u_4 + u_7] + (1-\alpha/2)[u_2 + u_5] \quad [4]$$

$$P_\infty(O_2 \cap A_1) = u_9 + \alpha u_{14} + (1-\alpha)u_{11} + [(1-\alpha)/2][u_{12} + u_{15}] + [1-(1-\alpha)/2][u_{10} + u_{13}] \quad [5]$$

Equation 1 gives the asymptotic probability of an O_1 response. Equations 2 and 3 present the asymptotic probability of an A_1 response on T_1 and T_2 trials, respectively. Finally, Equations 4 and 5 present the asymptotic probability of the joint occurrence of each observing response with an A_1 response.

METHOD

Experimental parameter values.—Six groups of Ss were tested. For all groups $\pi_1 = .9$. The groups differed with respect to the experimental parameters π_2 and α; three values of π_2 (.9, .5, and .1) and two values of α (1.00 and .75) were used. Specifically, $\pi_2 = .90$, $\alpha = 1.0$ (Group I); $\pi_2 = .50$, $\alpha = 1.0$ (Group II); $\pi_2 = .10$, $\alpha = 1.0$ (Group III); $\pi_2 = .90$, $\alpha = .75$ (Group IV); $\pi_2 = .50$, $\alpha = .75$ (Group V); and $\pi_2 = .10$, $\alpha = .75$ (Group VI). These particular values of π were selected because they had been used in a similar discrimination experiment where the observing response was not available (Atkinson, Bogartz, & Turner, 1959).

Subjects.—The Ss were 240 undergraduates obtained from introductory courses in psychology. They were randomly assigned to groups with the restriction of 40 Ss in each group.

Apparatus.—The Ss were run in subgroups of two with each S seated in a private booth. The apparatus, viewed from within S's booth, consisted of a shelf at table level which was 30 in. wide and 13 in. deep. A panel 30 in. wide and 30 in. high was mounted vertically on the edge of the shelf farthest from S. Four red panel lights (the s_i stimuli) were in a column and centered on the vertical panel; the bottom light was 20 in. from the base of the panel; the others were spaced above each other at 1¼-in. intervals. Two silent operating keys (the A_1 and A_2 responses) were each mounted 1½ in. in from the edge of the shelf facing S; these keys were 14 in. apart and centered on the column of red lights. On the shelf, 1 in. behind each of these keys, was a white panel light (E_1 and E_2 events). Two additional silent operating keys (the O_1 and O_2 responses) were each mounted 6 in. in from the rear edge of the shelf; these keys were 2 in. apart and also centered on the red lights. A green light (the signal) was centered 3 in. behind the observing response keys on the shelf. The presentation and duration of the lights were automatically controlled.

Procedure.—Within each of the six experimental groups, four subgroups of 10 Ss were formed by counterbalancing right and left positions of the observing response and the discrimination response keys. For each S one of the four red lights was randomly designated s_1, another s_c, and another s_2; the fourth light was not used.

The Ss were read the following instructions:

> The present study is designed to determine how well you can do on a very difficult pattern recognition problem. We run subjects in pairs to save time, but you are both working on completely different problems. The experiment for each of you consists of a series of trials. The green light on your panel will go on to indicate the start of each trial. Some time later, one or the other of the two lower white lights will go on. Your job is to predict on each trial which one of the two white lights will go on and to indicate your prediction by pressing one of the two lower keys.

> However, before you make your prediction you will receive additional information. That is, as soon as the green light goes on, press one or the other of the two upper keys—which key you press is up to you. Shortly thereafter, one of the four red lights will go on. The particular red light which goes on depends in part on the key you press. Further, the red light which goes on will help you in making your prediction as to which white light goes on. After you have seen one of the red lights go on, you will then predict which white light will go on by pressing the proper key. That is, if you expect the left white light to go on, press the left lower key, and if you expect the right white light to go on, press the right lower key. If the light above the key you pressed goes on, your prediction was correct, but if the light above the key opposite from the one you pressed goes on, you were incorrect and should have pressed the other key. Thus, for a single trial, the sequence of events is as follows: (1) the green light goes on to signal the start of the trial, (2) you press one of the two upper keys, (3) one of the red lights will go on, (4) you press one of the two lower keys, (5) if the white light goes on above the key you pressed, your prediction was correct; if the light above the key opposite from the one you pressed goes on, you were incorrect and should have pressed the other key.

Questions were answered by paraphrasing the appropriate part of the instructions. Following the instructions, 200 trials were run in continuous sequence. This sequence was followed by a 5-min. rest period; during this period no questions referring to the experiment were answered by *E*, and *S*s were not allowed to discuss the experiment. Following the rest, 200 additional trials were run. For each *S*, random sequences of s_i and E_i events were generated in accordance with assigned values of π_1, π_2, α, and the observed O_i responses.

On all trials, the signal light was lighted for 2 sec. The appropriate s_i stimulus light immediately followed the cessation of the signal light and remained on for 3 sec. After

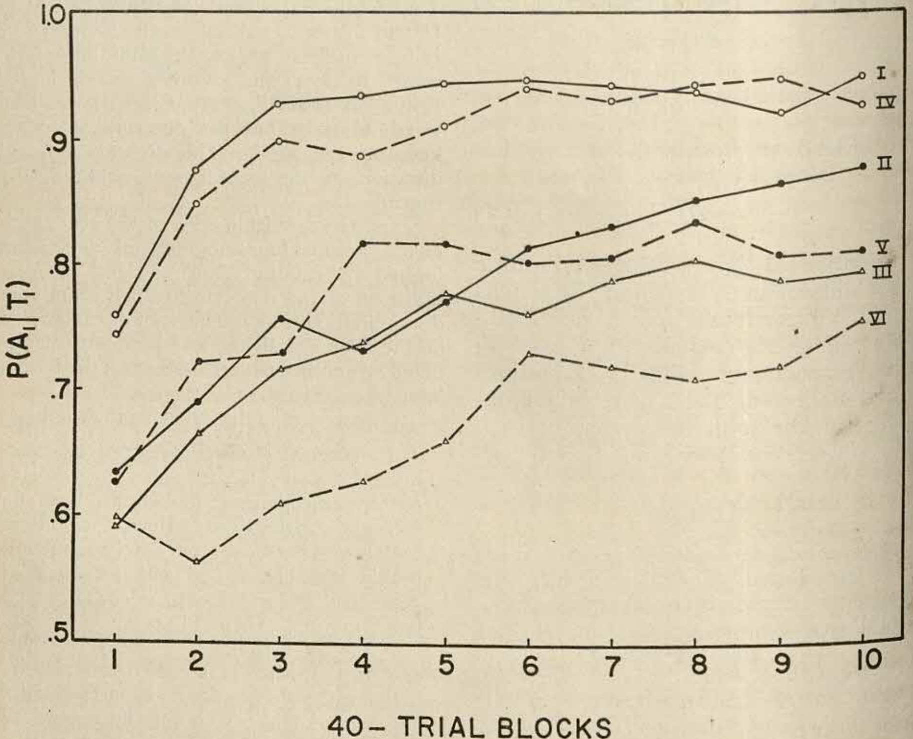


FIG. 1. The average proportion of A_1 responses on T_1-type trials in successive blocks of 40 trials.

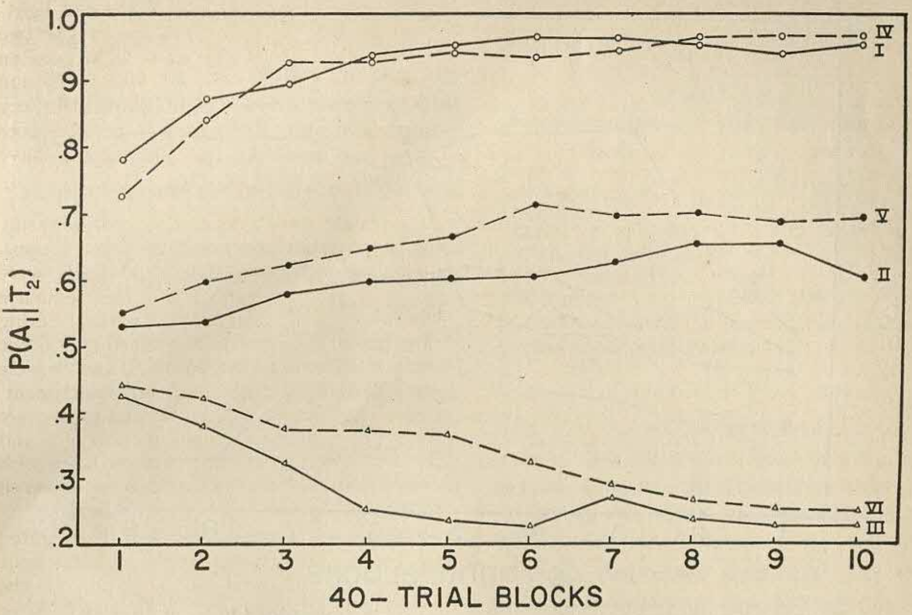


FIG. 2. The average proportion of A_1 responses on T_2-type trials in successive blocks of 40 trials.

the offset of the s_i light, one of the reinforcing lights went on for 2 sec. The time between the offset of the reinforcing light and the onset of the signal light for the next trial was 3 sec.

RESULTS AND DISCUSSION

Mean learning curves and asymptotic results.—Figure 1 presents the average proportion of A_1 responses on T_1-type trials in successive blocks of 40 trials. For each *S* the proportion of A_1's on T_1 trials was tabulated for a 40-trial block, and these quantities were then averaged over *S*s. Similarly, Fig. 2 presents the average proportion of A_1's on T_2-type trials in successive blocks of 40 trials. Finally, Fig. 3 presents the average proportion of O_1 responses. In all three figures the curves appear to be reasonably stable over the last half of the experiment. Consequently, the proportions computed over the final block of 160 trials were used as estimates of asymptotic performance.

Table 1 presents the observed mean proportions over the last 160-trial block and the related *SD*s. The observed values of $P_\infty(A_1 | T_i)$ were computed as indicated in the description of Fig. 1 and 2. The observed values of $P_\infty(O_i \cap A_1)$ were computed by obtaining, for individual *S*s, the proportion of trials on which both the A_1 response and the O_i response occurred and then averaging over *S*s.

The values predicted by the model are also presented in Table 1 for the case where $\theta' = \theta'' = \theta$. Expressions for the u_k's were derived by standard methods (Feller, 1957), and then combined by Equations 1–5 to predict the response probabilities. The computations were performed at the Western Data Processing Center on

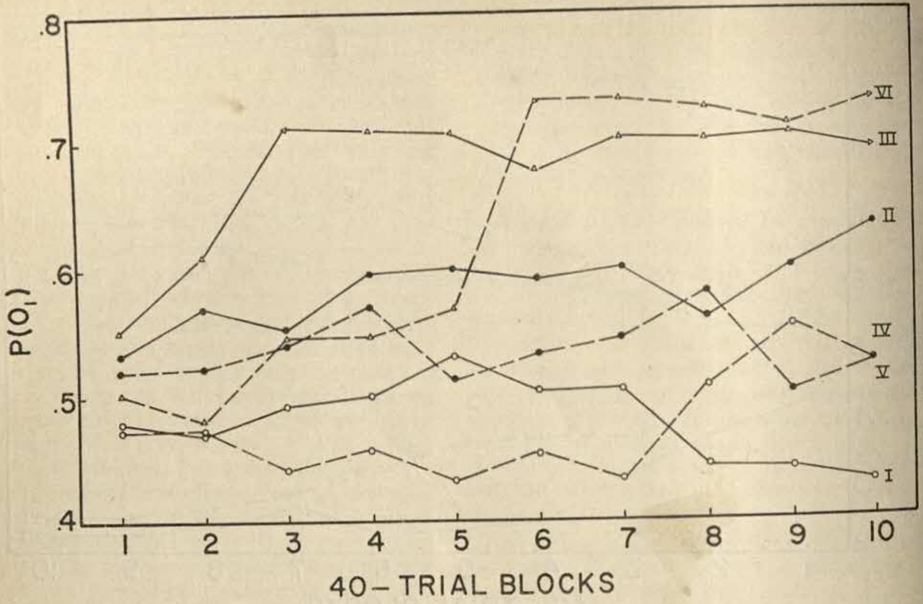


FIG. 3. The average proportion of O_1 responses in successive blocks of 40 trials.

TABLE 1

PREDICTED AND OBSERVED RESPONSE PROBABILITIES OVER THE LAST BLOCK OF 160 TRIALS

	Group I			Group II			Group III		
	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*
$P_\infty(A_1 \mid T_1)$	.90	.94	.014	.81	.85	.164	.79	.79	.158
$P_\infty(A_1 \mid T_2)$	.90	.94	.014	.59	.61	.134	.21	.23	.182
$P_\infty(O_1)$	.50	.45	.279	.55	.59	.279	.73	.70	.285
$P_\infty(O_1 \cap A_1)$	.45	.43	.266	.39	.42	.226	.37	.36	.164
$P_\infty(O_2 \cap A_1)$	.45	.47	.293	.31	.31	.232	.13	.16	.161
	Group IV			Group V			Group VI		
	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*
$P_\infty(A_1 \mid T_1)$	.90	.93	.063	.80	.82	.114	.73	.73	.138
$P_\infty(A_1 \mid T_2)$	.90	.95	.014	.60	.68	.114	.27	.25	.138
$P_\infty(O_1)$	.49	.50	.257	.52	.53	.305	.63	.72	.263
$P_\infty(O_1 \cap A_1)$	.44	.47	.241	.35	.38	.219	.32	.36	.138
$P_\infty(O_2 \cap A_1)$	.46	.47	.247	.34	.36	.272	.19	.13	.168

an IBM 709 computer.[2] By presenting a single value for each theoretical quantity in Table 1 we imply that these predicted proportions are independent of θ. Actually this is not always the case. However for the event schedules employed in this experiment the dependency of the theoretical proportions on θ is negligible. For θ ranging over the interval .00001 to 1.0 the values of the predicted proportions are affected in only the third or fourth decimal place; it is for this reason that we present theoretical values to only two decimal places.[3]

---

[2] The program or punch program deck is available to anyone interested in generating theoretical results for parameter values not considered in this paper.

In view of these comments it should be clear that the predictions in Table 1 are based solely on the experimentally assigned values of π_1, π_2, and α. Thus, they are entirely a priori and do not make use of any parameters evaluated from the data. Consequently, differences between *S*s, which can be represented by inter-*S* variability in θ, do not substantially affect these asymptotic predictions. Of course, this implies that the observed proportions for individual *S*s and also proportions averaged over *S*s should both approach these predicted values with increasing sample size.

An inspection of Table 1 indicates good agreement between observed and predicted quantities. The observed value of $P_\infty(A_1|T_1)$ decreases from Groups I to III and also from Groups IV to VI as predicted by the model. Similarly the observed values of $P_\infty(A_1|T_2)$ decrease from I to III and from IV to VI as expected. Column comparisons are also in the appropriate order, that is, on this measure Group I is less than IV, II is less than V, and III is less than VI. Thus, as we increase the frequency of reinforcing the A_2 response on T_2 trials, we not only observe an increment in $P_\infty(A_2|T_2)$ but also a decrement in $P_\infty(A_1|T_1)$.

For $P_\infty(O_1)$, an increase occurs from Groups I to III and from Groups IV to VI in accordance with theoretical results. That is, the proportion of O_1 responses increases as a function of the difference between π_1 and π_2; of course, this result would be expected in view of the fact that differential reinforcement for the observing responses depends on the difference between the reinforcement schedules on T_1 and T_2 trials. However, column comparisons on the $P_\infty(O_1)$ measure for I–IV and III–VI are in the reverse order; the difference on the $P_\infty(O_1)$ measure is particularly large for Group VI. This difference between data and theory for Group VI is also reflected in $P_\infty(O_2 \cap A_1)$; in fact, the discrepancies of these two quantities from predicted values are greater than any of the others in Table 1.

TABLE 2

ANALYSIS OF VARIANCE OF THE NUMBER OF O_1 RESPONSES IN THE LAST BLOCK OF 160 TRIALS

Source	*df*	*MS*	*F*
π (values of π_2)	2	27,533.2	13.8*
α values	1	113.4	0.1
O (O_1 left or right)	1	22,253.0	11.1*
A (A_1 left or right)	1	44.2	0.0
Interactions (11)	18		—[a]
Within	216	2,000.7	

[a] None significant at .05 level.
* $P < .001$.

An analysis of variance on the number of O_1 responses in the last block of 160 trials is presented in Table 2. The effects of the O_1 and A_1 placements on *S*'s panel (i.e., right or left) are included in the analysis. The effect of the π-variable is highly significant as would be expected. However, the α-variable is not significant. This finding might have been anticipated since the theoretical prediction for the over-all effect of α is small for the parameter values used in this study. The most unexpected result of the analysis is with regard to the observing response variable; the placement of the O_1 key turns out to be highly significant while the placement of the A_1 key has no effect. Over all groups and *S*s for the last 160 trials, the right hand observing response key was chosen on 55% of the trials while the right hand A_j key was selected on 50% of the trials. This right position preference on the observing response keys is particularly surprising in view of the fact that no similar preference exists for the A_j key. Several variables may account for this finding; for example, the observing response keys are in juxtaposition while the A_j keys are well separated; also, the observing response keys are further from *S* than the A_j keys.

---

[3] Essentially the same statement holds for $\theta' \neq \theta''$. However, in some cases the dependency is slightly larger.

In order to evaluate statistically the adequacy of the present model we have run a test suggested by Pillai and Ramachandran (1954) on the $P_\infty(O_1)$ measure. The test involves taking the largest absolute difference between an observed mean value and the predicted value in a collection of samples (in this case six). This difference is then divided by an over-all estimate of the standard error of the mean, that is, it is assumed that the observations are randomly selected from populations with homogeneous variance. As noted above, the largest discrepancy on the $P_\infty(O_1)$ measure occurs for Group VI. The predicted number of O_1s in the last block of 160 trials was 100.8 and the observed mean value was 115.7. The within-cells term in Table 2 was used to estimate the standard error of the mean, and in terms of Cochran's test there was no reason to reject the assumption of homogeneous variance. The obtained value of the Pillai-Ramachandran statistic was 2.1 and was not significant at the .05 level. Consequently, in terms of this particular statistical criterion there is no evidence to suggest that we reject the present model.

As noted earlier, not only group means but also the responses of individual *S*s should approach the theoretical values presented in Table 1. A check on the correspondence between individual asymptotic behavior and predicted values is equivalent to evaluating the agreement between observed *SD*s presented in Table 1 and asymptotic variability predicted by the model. Unfortunately direct computation of the theoretical *SD* is extremely cumbrous, and we have not obtained an analytical result. However, research reported by Suppes and Atkinson (1960) dealing with a similar model found that observed *SD*s were substantially larger than predicted values. Considering the rather large *SD*s reported here, their finding may be applicable to this set of data.

Transition characteristics.—A basic assumption in the model requires that if *S* is correct on trial n (i.e., A_1-E_1 or A_2-E_2 occurs) then on trial $n + 1$ he will repeat the observing response made on trial n. However, if *S* is incorrect (i.e., A_1-E_2 or A_2-E_1 occurs) then with probability θ'' he will shift observing responses from trial n to $n + 1$. This is a strong assumption and yields a highly deterministic set of predictions; for example, repetition of an observing response with probability 1 if *S* is correct on the preceding trial. On the other hand, a weaker form of the assumption which requires only a greater probability of observing response alternation following trials on which incorrect as compared to correct discrimination responses occur seems to be a reasonable conjecture for this type of problem. To test this class of assumptions we have computed the proportions of observing response alternations conditionalized on correct and incorrect discrimination responses over the last 160 trials. Let $N_n(s|c)$ denote the number of *S*s who were correct on trial $n - 1$ and shifted observing responses from trial $n - 1$ to n; also, let $N_n(c)$ be the number of *S*s who were correct on trial n. Similarly, define $N_n(s|\bar{c})$ and $N_n(\bar{c})$ in terms of incorrect responses. Further, define

$$N(s|c) = \sum_{n=241}^{400} N_n(s|c)$$

and

$$N(c) = \sum_{n=240}^{399} N_n(c)$$

and define $N(s|\bar{c})$ and $N(\bar{c})$ similarly. Then estimates of the conditional probabilities of shifting observing responses following correct or incorrect discrimination responses are, respectively,

$$\hat{P}(s|c) = \frac{N(s|c)}{N(c)},$$

$$\hat{P}(s|\bar{c}) = \frac{N(s|\bar{c})}{N(\bar{c})}$$

Table 3 presents the observed data for each of the groups. No statistical test is needed to see that these observed transition probabilities differ significantly from theoretical values. It suffices to note that theoretically $\hat{P}(s|c)$ should be identically zero for all groups whereas the observed values of this quantity differ markedly from zero. Without regard to the specific assumption considered in this paper, the question can be raised as to whether or not shifting of an observing response is more likely following incorrect or correct trials, that is whether $\hat{P}(s|\bar{c})$ is greater than $\hat{P}(s|c)$. A formal test of this hypothesis is a complex matter and we do not attempt it here. However note that for five of the six groups $\hat{P}(s|\bar{c})$ is greater than $\hat{P}(s|c)$. Further, the difference between these quantities increases as π_2 decreases; that is, the difference increases from Groups I to III and from Groups IV to VI.

TABLE 3

TRANSITION FREQUENCIES AND ESTIMATED PROPORTIONS OVER THE LAST BLOCK OF 160 TRIALS

	Groups					
	I	II	III	IV	V	VI
$N(s\|c)$	1610	880	933	1617	883	735
$N(c)$	5388	4080	4664	5463	4032	4350
$N(s\|\bar{c})$	275	520	446	266	517	524
$N(\bar{c})$	973	2280	1696	897	2328	2010
$\hat{P}(s\|c)$	.299	.216	.200	.296	.219	.169
$\hat{P}(s\|\bar{c})$	.283	.228	.263	.297	.222	.261

In conclusion, the rather striking correspondence between theoretical and observed values in Table 1 lends considerable support to the main features of the model. For the type of discrimination problem considered in this paper, it seems clear that asymptotic behavior can be predicted with accuracy in terms of the particular relations we have postulated among reinforcement schedules, observing responses, and discrimination responses. However, the sequential data reported in Table 3 indicate that some of the detailed features of the stimulus sampling process assumed in the model are certainly incorrect; this finding is not too surprising in view of related research on similar Markov chain models. Fortunately, within the framework of stimulus sampling theory, one can restate our axioms in only slightly modified form and thereby avoid the completely deterministic predictions made by the present model for sequential data. The disadvantage of such a reformulation is that the mathematical complexity of the model is greatly increased. The reader interested in details of such modifications is referred to Suppes and Atkinson (1960).

SUMMARY

An analysis of observing responses in discrimination learning was made. The typical discrimination task was modified so that two mutually exclusive and exhaustive observing responses could be identified and directly recorded. The experimental situa-

tion involved a series of 400 trials, each trial belonging to one of two types (T_1 or T_2). The sequence of events on a trial was as follows: (*a*) ready signal to which *S* made an observing response; (*b*) the presentation of one of three stimuli; (*c*) occurrence of one of two discrimination responses to the stimulus presentation; (*d*) termination of the trial with the reinforcement of a discrimination response. The particular stimulus presented on a trial depended on the observing response and the trial type. Following one of the observing responses, different stimuli were presented on T_1 and T_2 trials so that it was possible for *S* to identify the trial type; following the other observing response, the same stimulus was presented on both types of trials and hence *S* could not identify the trial type.

Six groups of college students were tested. The major independent variable specified different pairs of reinforcement schedules for the two trial types. The results indicated a highly predictable relation between the selection of observing responses and reinforcement schedules. In general, the greater the difference between the reinforcement schedules on T_1 and T_2 trials, the greater the preference for one observing response over the other. The analysis of the data was in terms of a Markov chain model which is closely related to stimulus sampling theory. There was excellent agreement between theoretical and observed values on asymptotic measures of observing and discrimination responses. However, an analysis of the sequential data indicated certain difficulties with the model.

REFERENCES

ATKINSON, R. C. A Markov model for discrimination learning. *Psychometrika*, 1958, **23**, 309–322.

ATKINSON, R. C. A theory of stimulus discrimination learning. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *First Stanford symposium on mathematical methods in the social sciences.* Stanford: Stanford Univer. Press, 1960. Ch. 15.

ATKINSON, R. C., BOGARTZ, W. H., & TURNER, R. N. Supplementary Report: Discrimination learning with probabilistic reinforcement schedules. *J. exp. Psychol.*, 1959, **57**, 349–350.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

FELLER, W. *An introduction to probability theory and its applications.* (2nd ed.) New York: Wiley, 1957.

PILLAI, K. C. S., & RAMACHANDRAN, K. V. On the distribution of the ratio of the i^{th} observation in an ordered sample from a normal population to an independent estimate of the standard deviation. *Ann. math. Statist.*, 1954, **25**, 565–572.

RESTLE, F. A survey and classification of learning models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Ch. 20.

SUPPES, P., & ATKINSON, R. C. *Markov learning models for multiperson interactions.* Stanford: Stanford Univer. Press, 1960.

WYCKOFF, L. B., JR. The role of observing responses in discrimination behavior. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received July 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 263–271

INTERACTION OF LACK OF SLEEP WITH KNOWLEDGE OF RESULTS, REPEATED TESTING, AND INDIVIDUAL DIFFERENCES[1]

ROBERT T. WILKINSON

Applied Psychology Research Unit, Cambridge, England

It has been established (Pepler, 1959; Wilkinson, 1958, 1959) that 30 hr. loss of sleep causes a decline in efficiency towards the end of 25 min. continuous work on Leonard's 5-Choice Test of Serial Reaction, a task in which a series of repeated simple choices have to be made as quickly and accurately as possible. These studies also suggested that this effect of lack of sleep changes little when the environmental warmth is increased and when periodic rest pauses are introduced. These negative results tell us a little about the nature and action of the stress but, perhaps, much more might be learned if we could find factors which influence its effect positively, either increasing or decreasing it. The present experiment has continued the search by examining whether the effect of lack of sleep changes, firstly as we increase feedback of knowledge of results (KR) in the test and, secondly, as *S* becomes more familiar with the test and with the stress. A final question was whether *S*s are consistent in the degree to which they are affected by lack of sleep. The two last-named issues were examined by having each *S* repeat an original 2-wk. program of testing twice so that their tests covered a period of 6 wk. Performances after sleep and no sleep on both an individual and a group basis could then be compared over successive, identical 2-wk. periods.

METHOD

Subjects.—The *S*s were 12 enlisted men between the ages of 18 and 30 yr. All were volunteers to undergo sleep deprivation. They were quartered near the laboratory for 6 wk. to act as *S*s in this experiment, and during this period they lived and worked together as a close social group.

Apparatus and procedure.—The apparatus and form of the 5-Choice Test of Serial Reaction has been described fully by Leonard (1959) and Wilkinson (1958). Briefly, there are five metal contacts and five associated bulbs. At the start of the test one of the bulbs is lighted and *S* taps the corresponding disc with a stylus; this causes the bulb to go out and another to light up, whereupon *S* again taps the appropriate disc, and so the cycle continues. The order in which the bulbs light up over successive taps is prearranged to appear random, and the task is merely to go on tapping the contact whose light comes on as quickly and as accurately as possible for 30 min. Scores are registered on three counters whose readings are recorded by *E* at the end of each 5-min. period of the test. One counter gives the number of correct taps made, another counter records an error whenever *S* taps the wrong disc; a third counts the number of gaps or occasions when *S* allows a period of 1.5 sec. or more to elapse between successive taps.

The variable of KR.—The test can take two forms, one with very complete feedback of KR and one where this feedback is at a minimum. In the former, *S* hears a bellnote whenever he makes an error and a buzz as soon as 1.5 sec. have elapsed since his last response, i.e., when he has scored a gap. Every 5 min. he hears *E*'s voice through a loud-speaker tell-

---

[1] The Royal Navy supported this work by contributing *S*s, administrative help, and financial assistance through the Royal Naval Personnel Research Committee of the Medical Research Council of the United Kingdom. The work owes much to the influence of N. H. Mackworth and D. E. Broadbent, successive Directors of the Applied Psychology Research Unit where it was carried out.

TABLE 1

Distribution of Conditions among Teams and over Days in Each of Three 2-Wk. Periods

Team	Day 1 (Tuesday)	Day 2 (Thursday)	Day 3 (Tuesday)	Day 4 (Thursday)
A (3 *S*s)	S:KR	NS:KR	NS:NKR	S:NKR
B (3 *S*s)	S:NKR	NS:NKR	NS:KR	S:KR
C (3 *S*s)	NS:KR	S:KR	S:NKR	NS:NKR
D (3 *S*s)	NS:NKR	S:NKR	S:KR	NS:KR

Note.—S:KR = sleep—knowledge of results; S:NKR = sleep—no knowledge of results; NS:KR = no sleep—knowledge of results; NS:NKR = no sleep—no knowledge of results.

ing him his scores of corrects and errors for the previous 5 min. work, and relating these to scores in the penultimate 5 min. by the comments, "Better," "Worse," or "No change." On the day following the test the full scores are published for all *S*s to see. In the other form of the test nothing is done to give *S*s any information on their performance. They hear *E*'s voice every 5 min., but he merely recites a short series of meaningless letters; about the same number of bells and buzzes are heard as in the comparable tests under full KR, but here they bear no relation to the incidence of errors and gaps and *S* knows this; finally, no results are published.

On the day preceding the first test *S*s were fully acquainted with the procedures described above and were also given 5 min. practice on each form of the test.

The variable of lack of sleep.—The method of sleep deprivation has been described fully elsewhere (Wilkinson, 1958). Tests after loss of sleep were always carried out at the same time of the day as the control tests. They differed from the latter only in that they employed *S*s who had foregone normal sleep the night before; these *S*s, staying awake in groups of three, were allowed to amuse themselves, under supervision, in their recreation room. No tests were given during the night.

Design.—The tests were carried out in four conditions: with KR after normal sleep, with KR after no sleep the previous night, with no KR after normal sleep, and with no KR after no sleep. The *S*s were tested on the Tuesday and Thursday of six consecutive weeks and by the end of the first 2 wk. each one had met each of the four conditions once in an order prescribed by his team as shown in Table 1. The second and third 2 wk. of the 6-wk. program were for each *S* an identical repetition of the first.

Results

Figure 1 is a comprehensive representation of the results, averaged over all 12 *S*s, in terms of gaps, corrects, and errors. In each of these categories three graphs are shown, one for each 2-wk. period; these show the average score in the four conditions over each 5-min. period of the test.

No sleep and no KR.—Figure 1 shows that these main effects were large and that they varied along the time-base of individual tests and along that of the 6-wk. program of tests. While this influence of time is important it may be simpler to ignore it for the present and assess these main effects on the basis of the total scores in the four conditions over the whole 6 wk. of testing. During this period each *S* met each combination of sleep or no sleep with KR or no KR three times. The difference between his scores in the sleep and no sleep tests will give an index of what we will call his "no-sleep impairment" or the degree to which his performance was impaired as a result of losing sleep. Dealing similarly with the tests with and without KR we can obtain an index of a "no KR impairment." These indices are shown in Table 2 derived from total scores of gaps, corrects, and errors in the four condition-combinations. Their significance is assessed by the nonparametric procedure of Wilcoxon (1949). Table 2 also shows the interaction of lack of sleep with no KR. This is obtained by taking the difference between no-sleep impairments in the tests with and without KR, the significance being assessed as before.

Lack of sleep caused a reduction in corrects ($P < .01$) and an increase in gaps ($P < .01$). Errors were increased also but not significantly.

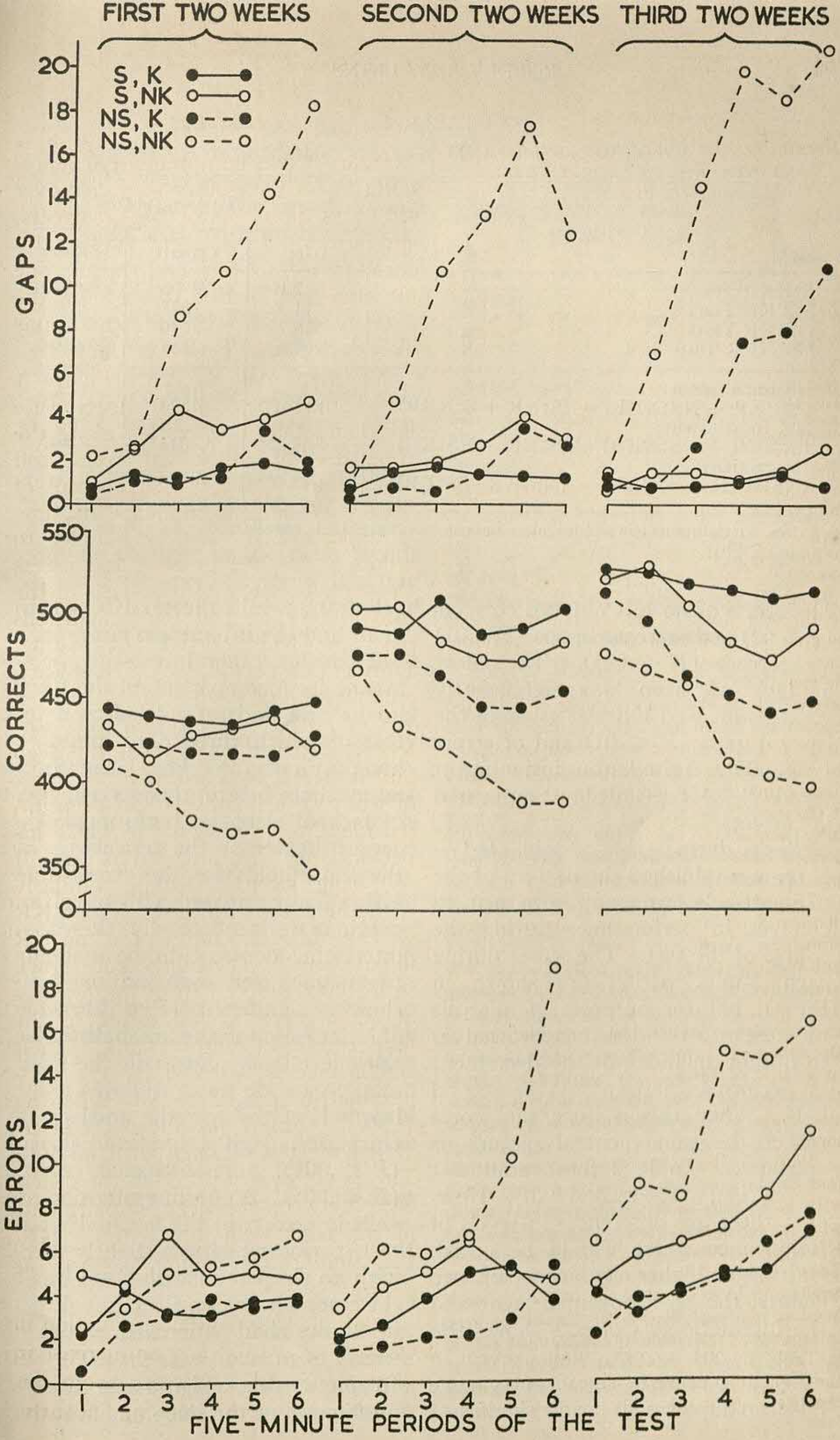


FIG. 1. The effect of combination of sleep and no-sleep with KR and no KR on gaps, corrects, and errors in three successive and identical 2-wk. periods of testing.

TABLE 2

EFFECT OF NO SLEEP (NS), KR, AND THEIR INTERACTION

Conditions	Scores Per Test Averaged Over All *Ss*		
	Gaps	Corrects	Errors
S:KR Tests	7.1	2906	22.6
S:NKR Tests	14.3	2827	33.5
NS:KR Tests	16.1	2696	20.1
NS:NKR Tests	66.3	2464	49.3
NS Impairment: (NS:KR + NS:NKR) − (S:KR + S:NKR)	+61.0**	−573**	+13.3
NK Impairment: (S:NKR + NS:NKR) − (S:KR + NS:KR)	+57.4**	−311*	+40.1**
Interaction: (NS:NKR − S:NKR) − (NS:KR − S:KR)	+43.0**	−153	+18.3*

Note.—Impairments and positive interactions appear as negative in corrects.
* $P < .05$.
** $P < .01$.

The stress of no KR reduced corrects ($P < .02$) and increased gaps ($P < .01$) and errors ($P < .001$). The effect of lack of sleep was significantly greater when no KR was given in the case of gaps ($P < .01$) and of errors ($P < .05$). A similar but insignificant trend ($P < .1$, one-tailed test) appeared in corrects.

Time during the 30-min. test.—Having established the pattern of the main trends, we can now turn to changes in performance during the course of the test. The scores during each 5-min. period are shown in Fig. 1, but for purposes of analysis the test was divided into two halves to obtain an index of the decrement in performance from first to second half. When this is done the scores reflect the same general picture as appears in Table 2 for overall test scores (i.e., first+second half). However there are differences in levels of significance which should be noted. Gaps show higher significance on two counts, the no-sleep impairment and the interaction. Corrects show little change, but errors again reveal a different picture of significance. The no-sleep impairment is now significant ($P < .02$) while the no KR impairment and the interaction cease to be.

There are other interesting points arising from comparison of the trends in the first and second halves of the test of which three only are noted here. Firstly, as has been shown many times before, the adverse effect of lack of sleep on performance becomes greater as the test wears on; the same holds for the stress of no KR, a finding not so well known from previous work. Secondly, there is an interesting contrast to be made between gaps and corrects; the latter show a significant ($P < .01$) effect of lack of sleep in the first half (indeed the effect is significant in the first 5 min.); gaps do not. And yet by the last half of the test the no-sleep impairment is highly significant in gaps ($P < .001$), more so than in corrects ($P < .005$). A puzzling situation prevails in errors; in the first half of the test, performance was slightly better with no sleep than with sleep but by the second half a no-sleep impairment in errors had appeared which is nearly significant at the .05 level. In spite of this, a significant ($P < .02$) interaction of the sleep and KR variables occurred in the first half which was reduced to insignificance in the last half. Presumably loss of sleep increases errors in the first half only when associated with no KR; by the last half the effect has spread to the tests where KR was given.

Time during the program of tests.—Changes during the course of the 12 tests performed by each S over a 6-wk. period will be considered only as they result from lack of sleep. Readers interested in variation of the no KR impairment and the interaction can distinguish these from the graphs in Fig. 1. Similarly, they may examine changes in the main trends across 2-wk. periods as a function of the deterioration in performance during the course of the test.

The no-sleep impairment increased from first to third 2 wk., significantly in the case of errors ($P < .005$) and of gaps ($P < .05$), but insignificantly in corrects, only 8 out of 12 *S*s showing the effect. The Wilcoxon test was again used to assess these levels of significance on a two-tailed basis; this time it was based on the difference in the no-sleep impairments occurring in the first and third 2 wk. These trends are shown in Fig. 2. It will be seen that the main increase has occurred by the second 2 wk. in corrects and errors but not until the third in the case of gaps. Errors were actually fewer after no sleep in the first 2 wk. Similar temporal trends were apparent during the course of the individual tests.

Consistency of Ss *in no-sleep impairment.*—The no-sleep impairment varied greatly between *S*s. If this index is used as a basis for ranking *S*s in each 2-wk. period then these rankings differ little from one 2 wk. to the next; those men who showed a large effect of lack of sleep in the first did the same in the second and third. This effect was significant in errors ($P<.05$) and corrects ($P<.05$) but not in gaps ($P < .2$)[2]. This consistency suggests that there are indeed real differences in the extent to which individuals are affected by lack of sleep.

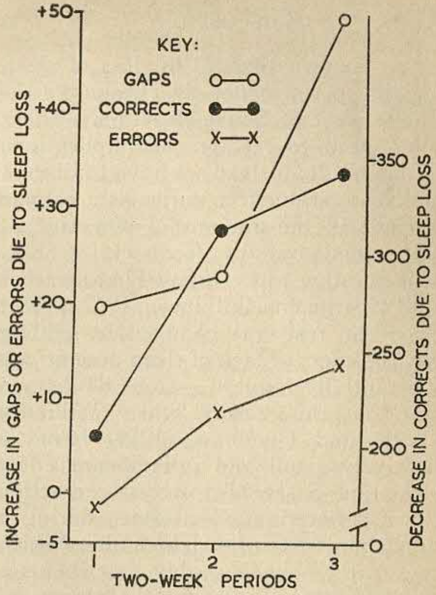


FIG. 2. Change in the impairment of performance due to lack of sleep over three successive and identical 2-wk. periods.

---

[2] The significance of these results was assessed by calculating Kendall's coefficient of concordance of the rankings in each team of 3 *S*s across 2-wk. periods. The method is described by Siegel (1956, p. 229). Probability values of the coefficients of concordance for each team were then combined to give an overall level of significance of the concordance of the rankings of all 12 *S*s across 2-wk. periods; this was thus independent of any concordance between teams due to their being consistently treated differently, i.e., with respect to the order in which they were given the four conditions (see Table 1). This is a relatively insensitive measure of significance being based only on 8 *df*; that .05 levels of significance were reached in spite of this encourages the view that in an experiment designed solely to investigate this question the same results would appear at much higher levels of significance.

Discussion

It is known that 30 hr. loss of sleep impairs performance in Leonard's 5-Choice Test of Serial Reaction. There is no need to restate the implications of this result; instead we have to discuss firstly how the effect varies as a change is made in the amount of incentive as determined by the feedback of KR. This question was examined in an earlier pilot experiment (Wilkinson, 1957, 1958); when the test was changed by adding KR the effect of lack of sleep upon it was substantially reduced. Some doubt was cast on this result when Williams, Lubin, and Goodnow (1959) found a relatively small and insignificant effect of letting *S* see his successive reaction times registering on a timer during a very simple test of serial reaction (which they rather challengeably describe as a vigilance task). One of the objects of the present experiment therefore was to check our earlier conclusion; it has been confirmed at a high level of confidence. The failure of the result of Williams et al. (1959) to reach significance may be due to the moderate nature of their feedback of KR, a possibility they themselves recognize.

We know that KR tends to raise the level of performance (Elwell & Grindley, 1938) and one reason for this is thought to be that it raises the level of motivation. Now we have seen that this effect is disproportionately large if *S*s have lost sleep and this suggests that in this state the level of motivation is unduly low, giving KR, so to speak, more to work on. Thus we now have some experimental evidence for the suggestion often made that lack of sleep produces a form of fatigue which does not directly cause a work decrement but raises the threshold at which certain work motives are effective (Whiting & English, 1925).

But what is motivation? A series of papers most of which have been mentioned in reviews by Malmo (1957, 1959), have sought to wed it to the concept of arousal. Although it must be conceded that the courtship has had its setbacks there certainly seems to be some relation between the level of activation or arousal of the body and level of performance (Bartoshuk, 1955; Surwillo, 1956), and with motivational variables which may influence this, for example, KR (Stennett, 1957), electric shock (Schnore, 1959), and interest content of a piece of work (Bartoshuk, 1956; Wallerstein, 1954). The present result provides further evidence for this by showing that a motivating influence (KR) can cancel the effects of an influence which according to the EEG findings of Bjerner (1949) and Armington and Mitnick (1959) makes for lowered arousal, namely lack of sleep. An important proviso to this is probably that the whole transaction must take place in the relatively low region of the arousal continuum, that is along the ascending arm of the inverted-U function which Duffy (1934), Freeman (1948), and Hebb (1955) have postulated as a link between activation and performance; in other words the test situation must be uninteresting or *S* very sleepy.

The interaction of two agents which are known to influence respectively motivation and arousal thus supports the idea of a close relation between the two concepts and may even suggest a common site of action neurologically, the obvious choice at our present state of knowledge being the brain stem reticular formation.

The second question raised by the present study concerns the influence of familiarity with the task and with the experience of going without sleep. Over a period of 6 wk. the present *S*s were repeatedly exposed to the same task and the same loss of one night's sleep. To the present author's knowledge this has been done only once before (Lee & Kleitman, 1923) but only one *S* was used and the outcome was inconclusive. The present results show at moderate but acceptable levels of significance that the influence of this familiarity was to heighten the effect of lack of sleep on the task. Three explanations may be considered: the first is that losing one night's sleep a week for 6 wk. has a cumulative effect. This seems unlikely; present *S*s had a whole week to recover from each night awake and the admittedly limited evidence we have suggests that one or two nights' normal sleep will suffice to remove the effects of a term of sleep-deprivation much longer than this. The second possibility is that after *S*s have foregone a night's sleep on six separate occasions the experience loses its novelty; *S*s may no longer feel challenged to make the extra effort which has often been held responsible for the failure of lack of sleep to affect performance (Edwards, 1941; Freeman, 1932; Robinson & Hermann, 1922). In the present experiment the decline of this stimulus towards the end of the 6-wk. program may have produced a more realistic and radical picture of the effect of lack of sleep on performance. Finally, having carried out the test twice a week for 6 wk., the test, rather than the experience of losing sleep, may have lost its novelty. It may be that one effect of lack of sleep is to increase the tendency for attention to be distracted from the main task by extraneous factors, especially when the task is one which does not compete well for attention. Berlyne (1950) has shown that decreased novelty can reduce a task's ability to compete in this way and this might explain why 6-wk. familiarity with the test increases the effect of lack of sleep upon it. Perhaps an important related point here is that the final state of decreased task novelty may also be a more realistic one in that many everyday tasks have reached at least a comparable state of familiarity. The lesson here may be that if we wish to assess the effect of lack of sleep and, indeed, that of other related stresses, for example noise and heat, on such everyday activities we may obtain more realistic answers from *S*s who, as in the present experiment, have been made thoroughly familiar through practice with the stress.

Before leaving this matter of the interaction of time with lack of sleep let us consider what happens during the individual test. It is well known that the effect of lack of sleep on performance can become greater with time spent continuously on the test, and we have seen this happen in the present experiment. But as a similar effect has appeared over the more protracted period of 6-wk. intermittent testing we can now compare the patterns of these two interactions. Interesting similarities exist. A no-sleep impairment in corrects has become noticeable by the second 5-min. period of the test and by the second 2 wk. of the 6-wk. program. In gaps this does not appear until the third 5 min. of test and the third 2 wk. of the program. Also errors show a unique pattern of negative impairment passing to positive impairment along both the 30-min. and the 6-wk. time scale. This correspondence suggests that the nature of the fatigue implied by the interaction may be the same whether time spent on the test is continuous or intermittent. This conclusion is in accord with that of an earlier study (Wilkinson, 1959) which discovered no difference in no-sleep impairment when a continuous test of 25 min. was compared with a similar one interspersed with 30-sec. breaks. The influence of both these results should be towards lessening the emphasis placed by Williams, Lubin, and Goodnow (1959) on the importance of *uninterrupted* work in contributing to a no-sleep impairment. It is tempting to conclude that the main casual factor in the interaction of lack of sleep with time on task, whether massed or spaced, is decreasing novelty. Our confidence in this interpretation is strengthened if we can assume that decreased novelty implies reduced reward from the activity of attending to the test, for then the conclusion reflects in a different context (the interaction of time with lack of sleep) the main finding of this experiment, namely that lowered reward (reduced KR) increased the effect of lack of sleep.

Wide individual differences exist in the extent to which lack of sleep impairs performance, and in studies of the subject it has been commonplace to note this. In the present experiment the consistency of these differences when assessed independently over three successive 2-wk. periods has shown them not to be due to chance.

This knowledge should encourage the search for measures, perhaps of personality, physique, or neurophysiological makeup, which correlate with them.

SUMMARY

The effect of one night's loss of sleep on a test of serial reaction was largely removed when the test was changed to allow greater feedback of knowledge of results. This interaction of lack of sleep with a known motivational factor suggests that the stress has one of its main effects in this area. To the extent that lack of sleep can be expected to lower arousal the result also supports suggestions of a close link between the concepts of arousal and motivation.

A basic 2-wk. program of testing was repeated three times on each *S* so that the effect of lack of sleep could be assessed independently in each of the three 2-wk. periods. The adverse effect on performance increased over the 6 wk.; three possible causes are considered: (*a*) a cumulative effect of loss of one night's sleep a week; (*b*) increasing familiarity with the stress; (*c*) increasing familiarity with the test. The two last-named explanations are preferred. It is suggested that the reduced novelty of both the stress and the test which prevailed during the last 2 wk. represents a more realistic situation; similar "acclimatisation" to the experimental situation might be of value in the examination of the effects of other stresses, for example, noise and alcohol, where factors of motivation and novelty may be important.

A further finding was that in the three identical and independent assessments of the effect of lack of sleep on the same *S*s, it was generally the same men whose performance was most impaired. Real individual differences, often suspected, are now confirmed and should be examined more closely.

REFERENCES

ARMINGTON, J. C., & MITNICK, L. L. Electroencephalogram and sleep deprivation. *J. appl. Physiol.*, 1959, **14**, 247–250.

BARTOSHUK, A. K. Electromyographic gradients as indicants of motivation. *Canad. J. Psychol.*, 1955, **9**, 215–230.

BARTOSHUK, A. K. EMG gradients and EEG amplitude during motivated listening. *Canad. J. Psychol.*, 1956, **10**, 156–164.

BERLYNE, D. E. Novelty and curiosity as determinants of behaviour. *Brit. J. Psychol.*, 1950, **41**, 68–80.

BJERNER, B. Alpha depression and lowered pulse rate during delayed actions in a serial reaction test. *Acta physiol. Scand.*, 1949, **19**, Suppl. 65.

DUFFY, E. Emotion: An example of the need for reorientation in psychology. *Psychol. Rev.*, 1934, **41**, 184–198.

EDWARDS, A. S. Effects of the loss of 100 hours of sleep. *Amer. J. Psychol.*, 1941, **54**, 80–91.

ELWELL, J. L., & GRINDLEY, G. C. The effect of knowledge of results on learning and performance. *Brit. J. Psychol.*, 1938, **29**, 39–53.

FREEMAN, G. L. Compensatory reinforcements of muscular tension subsequent to sleep loss. *J. exp. Psychol.*, 1932, **15**, 267–283.

FREEMAN, G. L. *The energetics of human behavior.* Ithaca: Cornell Univer. Press, 1948.

HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

LEE, M. A. M., & KLEITMAN, N. Studies on the physiology of sleep: II. Attempts to demonstrate functional changes in the nervous system during experimental insomnia. *Amer. J. Physiol.*, 1923, **57**, 141–151.

LEONARD, J. A. 5 Choice serial reaction apparatus. *Med. Res. Council Appl. Psychol. Res. Unit Rep.*, 1959, No. 326/59.

MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–287.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

PEPLER, R. D. Warmth and lack of sleep: Accuracy or activity reduced. *J. comp. physiol. Psychol.*, 1959, **52**, 446–450.

ROBINSON, E. S., & HERMANN, S. O. Effects of loss of sleep. *J. exp. Psychol.*, 1922, **5**, 19–32.

SCHNORE, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *J. exp. Psychol.*, 1959, **58**, 117–128.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

STENNETT, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61.

SURWILLO, W. W. Psychological factors in muscle-action potentials: EMG gradients. *J. exp. Psychol.*, 1956, **52**, 263–272.

WALLERSTEIN, H. An electromyographic study of attentive listening. *Canad. J. Psychol.*, 1954, **8**, 228–238.

WHITING, H. F., & ENGLISH, H. B. Fatigue tests and incentives. *J. exp. Psychol.*, 1925, **8**, 33–49.

WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid, 1949.

WILKINSON, R. T. The effects of lack of sleep on perception and skill. Unpublished doctoral dissertation, University of Cambridge, 1957.

WILKINSON, R. T. The effects of sleep loss on performance. *Med. Res. Council Appl. Psychol., Res. Unit Rep.*, 1958, No. 323/58.

WILKINSON, R. T. Rest pauses in a task affected by lack of sleep. *Ergonomics*, 1959, **2**, 373–380.

WILLIAMS, H. L., LUBIN, A., & GOODNOW, J. J. Impaired performance with acute sleep loss. *Psychol. Monogr.*, 1959, **73**(14, Whole No. 484).

(Received July 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 272–282

FIGURE COHERENCE IN THE KINETIC DEPTH EFFECT

BERT F. GREEN, JR.

Lincoln Laboratory,[1] *Massachusetts Institute of Technology*

When an observer views the two-dimensional (2-D) projection, e.g., shadow, of a moving three-dimensional (3-D) object, he usually perceives the shadow pattern as a form with depth. This has been called the Kinetic Depth Effect (KDE). Wallach and O'Connell (1953) concluded that an essential condition for the occurrence of the KDE seemed to be contours or lines that change their direction and their length simultaneously. Wallach, O'Connell, and Neisser (1953) showed that experience with an unfamiliar figure undergoing the KDE led to later perception of depth in the stationary shadow of the object. Gibson and Gibson (1957) studied the apparent slant of a 2-D surface formed by a set of regular or irregular forms. They obtained accurate judgments of slant when these forms underwent the continuous perspective transformations associated with plane rotation. Their experiments could be considered instances of the KDE but the Gibsons emphasize the importance of perspective for the perception of rigid motion, while perspective was apparently not an important determinant in the studies of Wallach et al.

In the KDE the relative motion of the various parts of a 3-D figure provides information not only about depth but also about the shape of the figure. The visual system in some way transforms the relative motions of the shadow's elements into the perception of a single rigid object, and this process appears to be a major factor in our normal perception of three-dimensional objects in the world around us. Johansson (1958) states the case as follows: "Mathematical relationships in the continuously changing energy distribution on the retina may be substituted for the classical static retinal picture, as the source of information from the external world. The substitution is viewed as an application of Gibson's gradient theory" (p. 3). It follows that a set of isolated dots or unconnected lines undergoing the appropriate changes can be perceived as a coherent rigid figure. White and Mueser (1960) studied the accuracy of the perception of arrangements of elements in a KDE setting. They found that all of their figures were perceived as rigid configurations in 3-D, but that reproducing the exact spatial relationships among the elements was a difficult task.

This paper reports a series of experiments designed to isolate the effect of relative movement from all other cues to depth and coherence. The experiments explore some of the conditions under which the KDE imparts perceived rigidity (coherence[2]) to elementary patterns of dots and lines. The *O*s viewed the changing 2-D projection associated with a rotating rigid 3-D configuration of dots or straight lines, and judged the apparent rigidity and coherence of the configurations, i.e., the extent to which the elements appeared to maintain their relative positions in the configuration. Independent variables were the number of elements in the configuration, the amount of constraint on the placement of elements, the type of rotation, the type of figure, and the speed of rotation.

---

[1] Operated with support from the United States Army, Navy, and Air Force.

[2] Rigidity and coherence are treated as synonyms in this paper.

GENERAL PROCEDURE

The stimuli were 16-mm. motion pictures showing the 2-D projections of rotating 3-D figures and were made by an animation technique. Each frame of the film showed the figure's 2-D projection from a certain 3-D orientation. Successive frames showed projections from a succession of different orientations that corresponded with a particular rotation of the figure. A succession of 250 frames formed the film strip for a figure. Projected at a rate of 20 frames per sec., the strip gave a 12.5-sec. movie of a rotating figure. The films were produced frame-by-frame, by an automatic camera that photographed a display generated on a cathode-ray tube by a digital computer. (The M.I.T. Lincoln Laboratory's Memory Test Computer, which has since been dismantled, was used to generate the stimuli in Exp. I, II, and III. The Laboratory's I. B. M. 704 computer was used for Exp. IV, V, and VI.) The computer was programed to display the 2-D perspective projections of a set of isolated dots or straight lines whose 3-D coordinates were stored in the computer. After making the display, the computer actuated the camera, advancing the film by one frame. The computer then calculated new 3-D coordinates for all the points or lines according to specified rotation formulas and repeated the cycle, displaying the 2-D projection, advancing the film, and rotating the figure, until the specified number of frames had been photographed. The particular series of pictures to be made on any computer run was controlled by the initial parameters, which included the initial coordinates of the dots or lines, the type of rotation, and the parameters of the rotation formulas. The computer program, in FORTRAN language, is available from the author. It is appropriate for the IBM Type 704 and 709 computers with an on-line CRT display.

The regular figures that we used were symmetric about the origin, with the origin being the centroid of the figure. The random figures were samples from a population for which the origin was the centroid, on the average. Two types of rotation were used: spinning about a fixed axis through the origin, and tumbling about the origin. In the spinning rotation, the figures revolved about a fixed axis as the earth revolves about its polar axis. The axis itself was not shown—it was merely used to specify the direction of movement. (The fixed axis is the locus of points that do not move.) In addition to specifying the axis, the programer specified the angular velocity of the spin about the axis. Except in Exp. IV, where speed was varied, the speed was always 64°/sec for all figures. In the tumbling rotation, the origin of the 3-D coordinate system was the only point that remained fixed. From any one frame to the next the tumbling rotation amounted to a spin about an axis, but throughout the film strip, the orientation of the axis changed continuously. To tumble the figure about a fixed point (the origin of the coordinate system) the programer used Euler's formulas (Snyder & Sisam, 1941, p. 42) to specify the rates of change of three angles, representing components of rotation about each of the three coordinate axes in turn. The speed of the tumbling rotation, as measured by the spin about the moving axis, was not constant. Also the movement of the axis did not have a constant velocity. Nevertheless the average angular velocity could be controlled; it was equated to the constant velocity of the spinning rotations.

The 2-D projections included perspective, which was computed in terms of the ratio of nominal figure diameter to nominal viewing distance. In all our experiments this ratio was about $\frac{1}{10}$, which is so small that the stimuli would have been nearly the same had we used a parallel 2-D projection rather than a perspective projection. The *O* sat 9 ft. from the projected display, on which the figure diameter was about 1 ft., making the visual angle 6½°. On the projected display, dots and lines were white on a dark ground. All dots had the same brightness and size, as did all lines.

Each configuration was displayed, rotating, for about 12.5 sec. The *O* was informed truthfully that he was actually viewing the 2-D projection of a rigid 3-D configuration, but he was instructed to rate the configuration as it appeared to him, rather than as he knew it to be. The *O* was told to use a 5-point subjective scale of rigidity or coherence, on which he was to give a rating of 1 if all the elements maintained their relative positions in the configuration throughout the exposure, and a rating of 5 if elements appeared to be moving independently. Intermediate ratings were to be given according to the proportion of coherent elements, and the relative amount of time that coherence was perceived. The

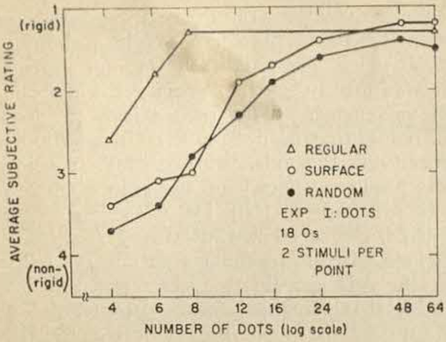


FIG. 1. Average ratings of rigidity in Exp. I, for three types of dot configurations. (In this and subsequent figures, the chance variability cannot easily be portrayed, because of the large individual differences. To assess the differences between any two figures, a difference in mean ratings of about $\frac{1}{2}$ rating step is roughly equivalent to a significant preponderance of ratings of one figure over another. When comparing points each based on two figures, the equivalent mean difference is about $\frac{1}{3}$ rating step.)

*O*s had no difficulty making the ratings after five practice trials. All *O*s were members of the laboratory staff, and were familiar with psychophysical experiments. Each *O* was tested individually, and made his responses orally.

EXP. I: CONSTRAINT AND NUMBER OF DOTS

Procedure

The first experiment used configurations of dots under three conditions of constraint. In the *random* condition, the 3-D coordinates of the dots were chosen so as to keep the dots within a 3-D cubical confine, i.e., each coordinate was chosen at random from the interval $-c \leq x \leq c$, where 2c is the length of one side of the confine. In the *surface* condition, dots were placed on the surface of a hypothetical cube or double pyramid (the latter has six vertices, $[0, 0, \pm c]$, $[0, \pm c, 0]$ and $[\pm c, 0, 0]$). In the case of the cube this was accomplished by setting one coordinate to $\pm c$, and choosing the other two at random. A similar method was used for the double pyramid. Each surface had the same number of dots as far as possible. For both surface and random conditions, eight values of n, the number of dots, were used: 4, 6, 8, 12, 16, 24, 48, and 64. In the *regular* condition, dots were placed at the vertices of a regular tetrahedron ($n = 4$), a regular double pyramid ($n = 6$) a cube ($n = 8$), and in a $4 \times 4 \times 4$ regular cubical array ($n = 64$). The tumbling rotation was used.

Eighteen *O*s were used. Two stimulus films were made for each of the 20 conditions (the regular conditions were repeated, but new random samples were drawn for the random and surface conditions). The 40 film strips were arranged at random in two blocks of 20 and spliced into a single film. Nine *O*s viewed the film projected in the forward direction, the others saw the film projected in reverse—they not only received the reverse order of stimuli, but also saw the stimuli moving in the reverse direction.

In this and subsequent experiments, the data will be reported in terms of the average ratings of coherence, pooling *O*s and stimuli within conditions. An attempt was made to scale the stimuli by the method of successive intervals. The results were in very close agreement with the simple averages of ratings, except for stimuli that most *O*s rated "1". For these, the scale values were very erratic, being violently affected by two or three ratings other than "1." Because of this instability, the scaling results will not be reported.

Results

The average ratings of coherence are shown in Fig. 1. Clearly, the regular configurations were judged to be more coherent than the surface and random configurations, for small n, while for 64 dots, all conditions yield the same high degree of perceived coherence. There is also a small but consistent difference between surface and random configurations, leading to the conclusion that the amount of constraint or regularity affects perceived coherence. The strong effect of the number of elements, n, is also clear from Fig. 1; the more elements, the more coherent a configuration appears.

An analysis of variance was made for the data from the random and surface constraints. The *O*s were divided into two groups, according to the order of presentation of stimuli,

TABLE 1

Analysis of Variance of Ratings of Coherence in Exp. I

Source	MS	df	Error Term	Error df	P	Variance Components
Number of dots (n)	6.4046	7	.3100[a]	12.2	.001	.7618
Groups (G)	2.9714	1	.1954	7	.01	.0868
Constraint (C)	.7855	1	.1415[b]	23	.05	.0201
NC	.0640	7	.1753	16	—	—
NG	.1954	7	.0607	16	.05	.0336
CG	.0123	1	.0607	16	—	—
NCG	.0955	7	.0607	16	—	—
NC(S)[c]	.1753	16	.0607	16	.05	.0573
NCG(S)	.0607	16	—			.0607

Note.—Procedures and notation follow Green and Tukey (1960).
[a] NG + NC(S) − NCG(S).
[b] NC(S) and NC pooled.
[c] Stimuli nested in NC.

giving an analysis of 2 constraints, 8 numbers of elements, 2 groups, and 2 stimuli nested in the combination of number of constraint. The analysis, shown in Table 1, indicated that both n and Groups have strong effects. The Group effect includes individual differences as well as presentation differences, and we suppose that the former predominates. The effect of constraint is barely significant at the .05 level. The significant n × Group interaction reflects the fact that the group differences, which in this analysis are surrogates of individual differences, are more apparent in the low values of n. For large n, most of the responses are "1." The analysis shows a large effect for n, and a strong effect of Group. The significant mean square for stimuli within n × Constraint indicates a small effect of particular stimuli.

Exp. II: Lines and Rotations

Procedure

The configurations in Exp. II were random line segments, the endpoints being specified in the same way as the dots in the random condition of Exp. I. The lines were either unconnected and independent, or connected, one to the next and the last to the first, to form a single closed 3-D curve. No 3-D curve lay in a plane—all had three dimensions. Two rotations were used: spinning about a vertical axis and tumbling. Five values of n, the number of lines, were used: 4, 6, 8, 12, and 16. Thus there were 20 conditions: 2 rotations × 2 types of figures × 5 values of n. Two stimulus movies were prepared for each condition, and were arranged in random order within each of two blocks of 20. Unfortunately the stimuli for the condition of 16 unconnected spinning lines were faulty and could not be replaced. Sixteen *O*s were used; most had served in Exp. I.

Results

The average ratings are shown in Fig. 2, averaged over *O*s and stimuli within conditions. Clearly the connected figures are seen as more coherent than the unconnected, and the spinning figures are seen as more

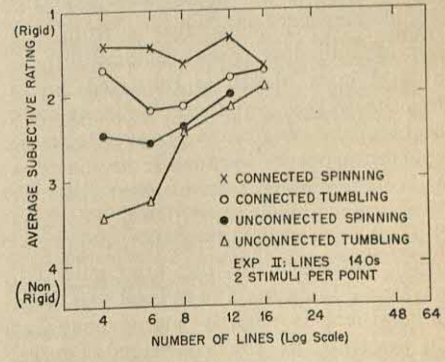


Fig. 2. Average ratings of rigidity for straight line figures in Exp. II.

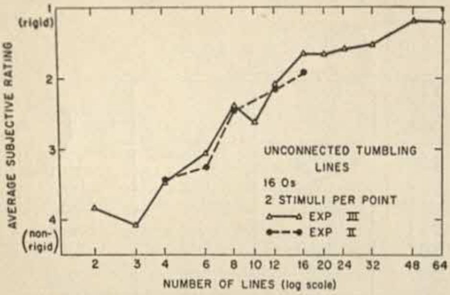


FIG. 3. Average ratings of rigidity for unconnected tumbling figures in Exp. III.

coherent than the tumbling. The statistical significance of these effects is established by a sign test. The effect of *n* is negligible for the connected lines, minor for the connected spinning figures, and important only for the unconnected tumbling figures.

The favorable effect of the spinning rotations is interesting. Rotation about a vertical axis is the most common kind of transformation seen when one walks about in the world. It is the type of rotation used exclusively by Wallach and O'Connell (1953). It is also a simpler rotation than tumbling, since every element moves at a constant angular velocity, and each element moves in a single plane, perpendicular to the axis of rotation. The questions of simplicity and familiarity are studied further in Exp. V and VI below.

The favorable effect of connected lines may be related to reversals. Since perspective is virtually absent from the stimuli, and since there are no brightness or size cues, depth is imparted solely by the KDE, so there is front-back ambiguity. In many cases, lack of coherence occurs because some, but not all, of the elements have reversed for *O*. With the unconnected lines, any one line can reverse independently of any other, but when the lines are connected, a reversal must involve at least two lines. Thus there is slightly more constraint on reversal in the connected figures, and this may account in part for the difference between the two types of figures. If the connected lines are interpreted as being more constrained than the unconnected, then the results of Exp. II are consistent with the effect of constraint found in Exp. I.

EXP. III: NUMBER OF UNCONNECTED LINES

Procedure

The third experiment extended the range of *n* in Exp. II for the unconnected tumbling figures, and also checked on the reliability of ratings. The stimuli for Exp. III were those from the unconnected tumbling condition of Exp. II, plus similar stimuli for additional values of *n* to cover the range of *n* from 2 to 64. In all there were two stimuli at each of 13 values of *n*, arranged at random in two blocks of 13 stimuli. The 16 *O*s used in Exp. II also served as *O*s in this experiment.

Results

The average ratings of coherence are shown in Fig. 3, together with the comparable curve from Fig. 2. It is clear that the ratings are adequately reliable, since the curves are about the same. Further, the effect of *n* for these figures is regular. Roughly, the average rating is linearly related to log *n*, at least up to 24 lines. Beyond 24, the average ratings decrease very little. The curve is very similar to the curve for the random dot figures in Fig. 1.

EXP. IV: SPEED OF ROTATION

Procedure

The fourth experiment was designed to study the effect of speed of rotation on judgment of coherence. Very slow and very fast speeds were expected to reduce the effectiveness of the KDE. Five speeds were used: 16, 32, 64, 128, and 256° per sec. The middle value, 64°/sec, was the speed used in all previous experiments. Five types of figures were used: 4, 16, or 64 random dots; 4 or 16 random unconnected lines. Only the tumbling type of rotation was used. Two versions of each speed × figure combination were prepared, yielding 50 stimuli. Each of 14 *O*s, most of whom had served in the previous experiments, saw the 50 stimuli in the same random order.

Results

The results are shown in Fig. 4. Clearly there is very little effect of speed over the range in our experiment. There is a hint of slightly less coherence at the slowest speed, but the other four speeds, covering an eightfold range, show no differential effects on coherence.

The range of speeds covered in this experiment nearly spans the range of possible speeds in our experimental procedure. Another factor of two faster would give so much change from one frame to the next (25°) that apparent movement would probably be destroyed. Another factor of two slower would result in a total excursion of only about 100° in the viewing time for each figure; no point in the figure could move much more than half way across the display, so that to get meaningful judgments of coherence, the viewing time would have to be increased. White and Mueser (1960) found a similar effect of speed on accuracy of perception in their KDE setting. They obtained differences among speeds of 2, 8, and 32 rpm in a spinning rotation, which correspond with 12°/sec, 48°/sec, and 192°/sec. They found reduced accuracy with a slower speed of 1 rpm (6°/sec), but they used an exposure time of 30 sec., and their results suggest that a longer viewing time would have allowed greater accuracy. Slow speed by itself may not detract from the KDE.

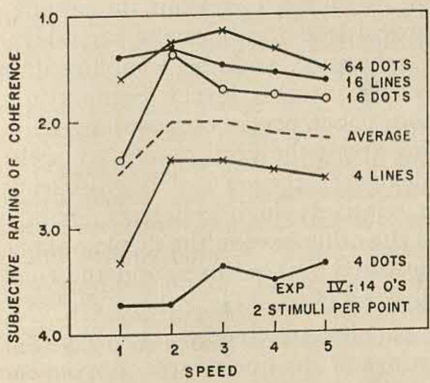


FIG. 4. Average ratings of rigidity for figures tumbling at five speeds in Exp. IV.

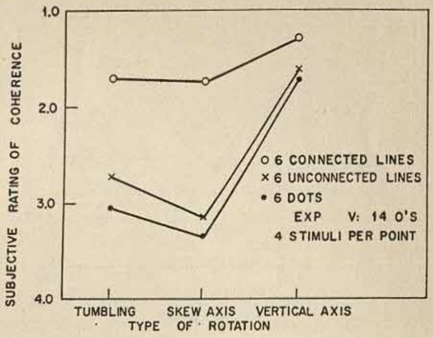


FIG. 5. Average ratings of rigidity for figures under three types of rotation in Exp. V.

EXP. V: SKEW SPINS

Procedure

A large difference was obtained in Exp. II between tumbling and vertically spinning. Experiment V was intended as a demonstration that spinning about a nonvertical axis would yield coherence judgments that were somewhere between those obtained with a vertical spin and with tumbling. The notion was that the judgments of coherence were related to the complexity of the rotation.

Three types of rotation were used: tumbling, spinning about a vertical axis, and spinning about a skew axis that was at an angle with all three coordinate axes. Two different skew axes were used. The cosines of the angles made by one skew axis with the horizontal axis in the display plane, the vertical axis, and the axis perpendicular to the display plane, were in the ratios 1:2:3, respectively. The other skew axis had cosines in the ratios 1:1:1. Three types of figures were used: six random dots, six unconnected random lines, and six connected random lines. Four versions of each figure type were made for the vertically spinning and the tumbling rotations, and two versions of each were made for each skew axis, yielding 36 stimuli. The 14 *O*s who had served in Exp. 4 were enlisted for Exp. 5.

Results

The results are shown in Fig. 5. Since the results for the two skew

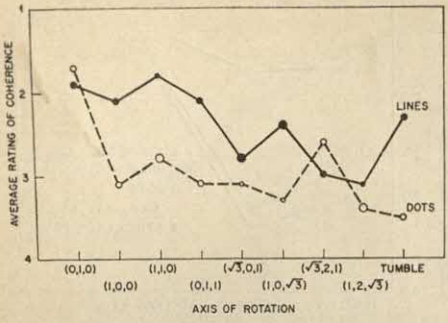


FIG. 6. Average ratings of coherence for various rotations in Exp. VI. (Small points represent 16 *O*s rating one stimulus; medium points, 16 *O*s rating 2 stimuli; large points, 16 *O*s rating 3 stimuli.)

axes did not differ more than differences between stimuli for a given axis, the results for the two skew axes were averaged. The results for the skew axes are in no sense intermediate between tumbling and vertically spinning. If anything, tumbling is intermediate. Clearly the effect of axis of rotation is more complicated than we had supposed. Another experiment was designed to study a variety of axes of rotation, in an attempt to resolve the complication presented by Exp. V.

EXP. VI: AXIS OF ROTATION[3]

Procedure

Nine patterns of rotation were used: the tumbling rotation and eight different fixed axes with the spinning rotation. The position of an axis is best specified by its direction numbers, which are numbers proportional to the cosines of the angles made by the axis with the three coordinate vectors: the horizontal vector in the display plane, the vertical vector, and the vector perpendicular to the display plane, respectively (Snyder & Sisam, 1941). The eight fixed axes of rotation used in Exp. VI can be divided into three groups. Group 1 includes axes in the plane of the display: (0, 1, 0), the vertical axis; (1, 0, 0), the horizontal axis in the display plane; and (1, 1, 0), an axis at an angle of 45° with both the horizontal and vertical vectors in the display plane. Group 2 includes axes not in the display plane but perpendicular to one reference axis: (0, 1, 1), an axis tilted forward 45° from the vertical and 45° up from the axis perpendicular to the display plane; ($\sqrt{3}$, 0, 1) and (1, 0, $\sqrt{3}$), axes 30° and 60° back from the display plane, respectively, in a plane perpendicular to the vertical reference axis. Group 3 contains two skew axes that made acute angles with all three reference vectors: ($\sqrt{3}$, 2, 1) and (1, 2, $\sqrt{3}$), axes 30° and 60° forward from the display plane, respectively, and both at 45° from the vertical reference vector.

Two types of figures were used; six random dots and six random unconnected lines. Two replications of each figure type were intended to be made for each type of rotation, yielding 36 conditions. After the experiment had been completed, it was discovered that for axes (1, 1, 0) and (1, 2, $\sqrt{3}$) there were three dot figures and one line figure, while for axes ($\sqrt{3}$, 0, 1) and (1, 0, $\sqrt{3}$) there were three line figures and one dot figure, instead of two each. Sixteen *O*s, most of whom had served in the previous experiments, rated the displays for coherence in the usual way.

After the ratings had been obtained from all *O*s, 13 *O*s were tested again. This time *O* was asked to indicate the direction of the axis of rotation by setting a rod in a matching position. The *O* was instructed to make a setting, even if there did not seem to be a single fixed axis of rotation.

Results

The results of the coherence ratings are shown in Fig. 6. For the dot figures, all fixed axes but the familiar vertical lead to about the same level of coherence, and the tumbling dots are also at this level. Significantly more coherence is obtained by spinning about the vertical axis. For the line figures, the vertical axis also leads to relatively high coherence, but so do the other axes in the display plane, the tilted vertical axis, and the tumbling rotation. The other skew axes impart less coherence. The coherence ratings of the line figures support the hypothesis that the more complex axis locations are associated with less coherence. The three groups of axes defined above were formed on a priori grounds to represent three levels of complexity. When the ratings for each *O* are averaged across all line figures within each group of axes, we find that for 11 of the 16 *O*s the group averages are in the order 1, 2, 3, a preponderance significant at the .0001 level. Very little support for the hypothesis can be obtained from the ratings of the dot figures, since only the vertical axis stands out.

---

[3] The author is indebted to Alice Wolf for running part of this experiment.

The results of the rod settings are shown in Fig. 7, where the proportion of setting within 15° and within 30° of the true axis are plotted for each axis. In order to plot the settings from the tumbling figures on the same figure, the median setting for the tumbling displays was taken as the "true" axis (the median was determined separately for dots and lines but the judgments from the two replications of each figure were pooled). The settings for the axes not in the display plane provide evidence for the reversals seen by *O*s. All of the true axes were in front of the display plane (the end of the axis above the origin of the display is used for this reference). Actually 35% of the settings had the axis in front, 50% in back, and 15% on the display plane. Thus the figures were seen at least as often with front and back reversed as in the programed orientation. For this reason, reversals were not counted as errors. There were two equally correct settings for all axes not in the display plane. On this account, the chance level for Group 1 differs from that of Groups 2 and 3. Chance levels are computed on the basis of a uniform distribution of settings over the surface of a unit sphere, and are shown in Fig. 7.

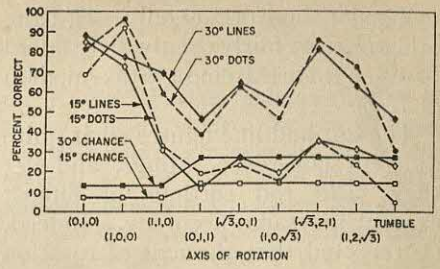


FIG. 7. Proportion of axis settings within specified angle of the true axis, in Exp. VI. (Two criteria of correctness are shown; 15° and 30°. Computed chance levels are shown for both criteria.)

Figure 7 shows that the settings for the tumbling figures were very inaccurate, in terms of the arbitrary "true" axes, indicating that the settings were quite variable. For the fixed axes, the stringent criterion of 15° error separates the vertical and horizontal axes from all the rest, and shows no differences among the rest. It also shows that setting the axis was a difficult task not accurately done by the *O*s. The sparse figures (only six elements in each), the short viewing time, and the tricky response device all contributed to the difficulty. Also, one or two *O*s occasionally set the rod at right angles to the true axis, apparently matching the direction of movement rather than the axis about which the movement occurred.

Measured by the liberal 30° criterion, the rod settings showed significant differences among the fixed axes, but no clear pattern emerges. Considering the three groups of axes described above, Groups 1 and 3 appear to have been matched with about the same accuracy but because of the differences in chance level, the settings for Group 1 would seem to be actually better. The axes in Group 2 were matched less accurately than the others. Individual differences in accuracy are huge. The best *O* had 97% of his settings within 30° of the true axis; the worst *O* only 19%; the other *O*s were fairly evenly distributed between these extremes, with a median of 78%.

The dot and line figures elicit quite different coherence ratings, and yet they yield rod settings of almost exactly the same accuracy. Indeed, *O*s reported that the axis of rotation was apparent in the dot displays even when the figures did not cohere. That is, even if some dots "broke away" perceptually from the figure, the direction in which the majority of the dots were rotating was still clear. In the case of dot figures rotating about axes in the display plane, about half the *O*s reported sometimes seeing the dots moving in 2-D, sliding back and forth on their straight paths (with sinusoidal velocity) rather than rotating in a 3-D circular path (with constant velocity). Nevertheless the proper setting for the rod was obvious.

There are other indications that matching the axis was psychologically quite different from judging coherence. The *r* between the average coherence rating of a figure and the number of errors made in matching its axis of rotation was only 0.16. The rank correlation between the number of ratings of "1" given by an individual and the accuracy of his settings was −.45.

DISCUSSION

It is abundantly clear that the KDE alone is a very powerful factor in depth perception. Many of our *O*s saw every figure in 3-D; some *O*s reported occasionally seeing figures in 2-D, but the occasions were rare and fleeting, even with the meager figures that were used. The tendency to see depth in the 2-D projections of rotating 3-D figures is overwhelming. There is also a strong tendency for points or line segments undergoing the same 3-D rotation to be seen as parts of a rigid, coherent figure. Even randomly placed, unconnected points and lines form a rigid figure if they undergo the same 3-D rotation, while connecting the elements, or otherwise constraining them enhances the tendency to perceive a rigid figure.

The KDE is powerful, but it is not a *sufficient* condition for perceiving complete coherence. The majority of our figures were perceived as 3-D but nonrigid. The "1" rating was used less than half the time, and the "5" rating was seldom used, indicating that *O*s judged most figures to be partially coherent. Matching the axis of rotation in Exp. VI was not done very accurately, despite the omnipresence of the KDE. White and Mueser (1960) found that the spatial arrangements of elements were not accurately perceived even though the configuration was perceived as rigid and three-dimensional. The KDE seems to be more immediate than perceived rigidity, and both seem more immediate than accuracy of perception of unfamiliar random figures.

The role of perspective in the KDE is not determined. Gibson and his coworkers (1957, 1959) emphasize the importance of changes in perspective transformations. In our experiments, on the other hand, perspective was essentially irrelevant. The *O*s reported that reversals were common, and the axis-matching data of Exp. VI show that there was no agreement about which was front and which was back. (In retrospect, it would have been better to use parallel projection in our experiments, to make our statements about perspective unequivocal.) The truth must be that perspective is effective when it has a chance to be; perspective is very important when it is essentially the only clue to depth, but it is no more an essential feature of depth perception than is any other single factor. Braunstein (1960)[4] has data that confirm this supposition.

Some of the characteristics of figures that influence perceived coherence have been established. The number of elements is of prime importance, as are connections and constraint among elements. From *O*s' reports we know that isolated elements have a tendency to "break away" from the main body of the figure, and that lines that cross each other seem to enhance coherence. But our experiments did not include curved lines, solid figures, or familiar forms. Much remains to be learned about the properties of figures that are related to the KDE.

---

[4] M. Braunstein, personal communication, 1960.

There is evidence of consistent individual differences across experiments. Twelve persons served as *O*s in all six experiments. For each experiment these *O*s were ordered according to the number of stimuli given a rating of 1. The similarity of the six resulting rankings of the 12 *O*s was evaluated by Kendall's coefficient of concordance, W. We found that $W = .60$, which is not only significantly different from zero at well beyond the .001 level, but also indicates very considerable consistency. The consistency of individual differences is all the more remarkable because the experiments were carried out at various times during a 3-yr. period. In each experiment, the percentage of "1" ratings given by the various *O*s varies over a wide range; the average range is 53 percentage points. How much of such large, consistent differences is due to differences in criteria for the "1" response, and how much to differences in perception cannot be determined from our data.

The subjective rating procedure served its purpose well. The ratings were reliable, and gave adequate measures of the major effects studied in these experiments. However, the distributions of ratings obtained in various conditions overlapped considerably, and we feel that a more sensitive measure is needed to investigate some of the finer points of the KDE. Also, individual differences in perception are confounded with individual differences in subjective scale values. To extract the individual differences which are known to be large, one must take enough data to establish firmly each *O*'s subjective scale, or else one must use an objective procedure.

How accurate is depth perception in the KDE? We may assume that when *O* saw a figure as rigid, he saw either a true 3-D figure or a complete front-back reversal of the true figure since there was only one rigid 3-D figure that would produce the 2-D display. But we have no other evidence that *O* saw the true figures, and indeed, the inaccuracy in the rod settings of Exp. VI suggests the possibility that depth is not perceived accurately. Accuracy of depth perception probably depends on the number of elements and the coherence of the figure, but the nature of this dependence is unknown.

There are many other unsolved problems. Can *O* discriminate between a rigid and a nonrigid 3-D figure, when perceived in the KDE situation? What happens with only one or two points or lines, i.e., what is the minimal stimulus for the KDE? Why are the more complex axes of rotation associated with less coherence? When is perspective important? Must all of the elements of the figure be continuously in view, as they were in these experiments? (Casual observation, plus everyday experience, strongly suggest that the answer is "No.") A wide field for research exists.

Finally, the role of the digital computer in this research deserves comment. Developing the computer techniques and camera techniques were difficult and time consuming, but without the computer, we would not have done the experiments at all. In our present setup, a film strip of 250 frames showing one figure rotating for 12.5 sec. can be produced in about 2 min., and there is complete assurance that all the elements in the figure will be of uniform size and brightness. Of course, KDE experiments can be done without computers, but the computer provides great flexibility in the kinds of experiments one can envisage and perform.

SUMMARY

Six experiments examined the extent to which the kinetic depth effect produces per-

ceived coherence and rigidity of random figures. Subjectively rated coherence was greater with (*a*) more elements in the figure, (*b*) connections and constraints among the elements, and (*c*) less complex axes of rotation. The tumbling rotation was shown to be intermediate between simple and complex axes of spinning rotations. Speed of rotation was shown to have almost no effect. Individual differences were large. A high speed digital computer was essential for producing the stimuli for the experiments.

REFERENCES

GIBSON, J. J., & GIBSON, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54**, 129–138.

GIBSON, E. J., GIBSON, J. J., SMITH, O. W., & FLOCK, H. Motion parallax as a determinant of perceived depth. *J. exp. Psychol.*, 1959, **58**, 40–51.

JOHANSSON, G. Rigidity, stability, and motion in perceptual space. *U. Uppsala Psychol. Lab. Rep.*, 1958, No. 3.

SNYDER, V., & SISAM, C. H. *Analytic geometry of space.* New York: Holt, 1941.

WALLACH, H., & O'CONNELL, D. N. The kinetic depth effect. *J. exp. Psychol.*, 1953, **45**, 205–217.

WALLACH, H., O'CONNELL, D. N., & NEISSER, U. The memory effect of visual perception of three-dimensional form. *J. exp. Psychol.*, 1953, **45**, 360–368.

WHITE, B. W., & MUESER, G. E. Accuracy in reconstructing the arrangement of elements generating kinetic depth displays. *J. exp. Psychol.*, 1960, **60**, 1–11.

(Received July 9, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 283–290

CONCEPT IDENTIFICATION USING SIMULTANEOUS AUDITORY AND VISUAL SIGNALS[1]

DANIEL S. LORDAHL[2]

University of Wisconsin

Among the variables affecting performance in concept identification tasks, the amount of irrelevant information in the stimuli has been most often studied. Archer, Bourne, and Brown (1955), Bourne (1957), Bourne and Pendleton (1958), and Brown and Archer (1956), have all varied the amount of irrelevant information in concept tasks. These studies differed with respect to other independent variables studied, but each showed that increases in irrelevant information caused significant increases in the difficulty of the task.

Bourne (1957), Bourne and Pendleton (1958), and Brown and Archer (1956) all found amount of irrelevant information (in bits) in problems of this type increases the difficulty of the task in a linear manner within the limits studied. Moreover, this effect holds over a number of minor variations in procedure. There were, however, a number of parameters in this set of experiments which, if varied, might upset the generality of the effect of irrelevant information. For one thing, each of the above studies used visual stimuli exclusively. A change of sense modality probably would not have changed the relationship between performance in concept identification and irrelevant information, but there is no assurance of this. Even less can be presumed about the effect of presenting information through two senses simultaneously. The present study investigates just this situation: the simultaneous presentation of information through the visual and auditory senses.

Some recent theoretical approaches have been made to the concept situation in terms of statistical learning theory (Bourne & Pendleton, 1958; Bourne & Restle, 1959; Walker, 1958). Although the present study was designed to answer the empirical problem indicated above, it also has implications for the mathematical theories of concept identification.

METHOD

Subjects.—The *S*s were volunteers drawn from the elementary course and certain intermediate courses in psychology at the University of Wisconsin during the spring and summer of 1959. The data from 96 *S*s are included in this paper. The data from four groups of 2 *S*s were dropped on account of procedural or apparatus difficulties.

Design.—The major variables under consideration were amount of irrelevant information presented visually and amount of irrelevant information presented auditorily. Under all conditions, one stimulus dimension from *each* mode of presentation was relevant to the solution of the problem presented to *S*s. Since only two levels of each dimension were used, in all conditions 1 visual and 1 auditory bit of information had to be taken into account by *S*s in order to solve the problem.

Three different problems were used, i.e., three different pairs of relevant stimulus dimensions (one visual and one auditory). It should be noted that visual and auditory stimuli were presented simultaneously.

---

[1] This paper is based on a dissertation submitted to the Graduate College of the University of Wisconsin in partial fulfillment of the requirements for the PhD degree. The writer is indebted to E. James Archer, who directed the study, for his aid and advice throughout the course of the experiment. The research was supported, in part, by funds from the Wisconsin Alumni Research Foundation.

[2] Now at Washington University, St. Louis, Missouri.

The design involved four amounts of irrelevant visual information, four amounts of irrelevant auditory information (0, 1, 2, or 3 bits in each case), and three problems. Two *S*s served in each of the 48 conditions, with each *S* receiving 192 trials.

The *S*s were required to make one of two lever-pressing responses for each of the 192 stimuli presented. All stimuli contained one of four possible combinations of the two bilevel relevant dimensions. One level from each relevant dimension was randomly designated as a *positive level.* When both positive levels were present in a given stimulus, that stimulus will be called a *positive instance* of the concept. All other stimuli will be called *negative instances* of the concept. When a positive instance of the concept was presented, the correct response for *S* was to press the positive lever, marked with a plus sign. The correct response for all negative instances was to depress the unmarked lever. These two responses, whether correct or incorrect, will be called *positive* and *negative lever responses*, respectively.

Apparatus.—The 2 *S*s in each cell served at the same time and were seated at a long table with an opaque panel separating them. Each *S* was supplied with a set of earphones (crystal headphones, Model BA-200B; Clevite Electronic Components, Cleveland, Ohio) through which the auditory stimuli were presented. On the table, directly in front of each *S*, a gray box (6×3×3 in.) was fixed. This box had two black-handled lever-action switches (3-in. separation) extending toward *S* from the front side, and above one of the switches there was a large black plus sign on top of the box. To the right of each switchbox was a second box approximately 10 in. high which had two 4 × 4 in. milk glass screens (one above the other) in the side facing *S*. These screens provided *S* with information about the appropriateness of his lever-pressing response; the top glass was illuminated from within the box with a white light for a correct response and the bottom glass was illuminated with a red light for an incorrect response.

Ten feet in front of the 2 *S*s was a gray partition separating *S*s from *E* and his equipment. In the partition in front of the 2 seated *S*s and about 1 ft. above eye level was a rectangular white translucent paper screen (12 × 10 in.) which was illuminated from behind by a strip film projector to provide the visual stimuli. The visual portion of the stimulus materials was photographed on 16 mm. film strips and projected on the screen from behind the partition.

The auditory signals were generated by an apparatus consisting basically of two audio oscillators and a white noise generator. Relay switching circuits varied the audio signal on five independent bileveled dimensions. The combinations of these dimensions in their different levels allowed for the possible presentation of 32 distinctly different auditory signals. The selection and presentation of the auditory signals was controlled by punched programs through Western Union Tape Transmitters.

The *S*s' responses were recorded graphically with an Esterline-Angus 20-channel operations recorder and, in addition, errors were automatically totaled for each *S* on counters.

The information lights in front of *S* were lighted immediately upon his lever-pressing response and remained on until the next stimulus was presented, 2 sec. following the response of the slower *S*. Also present at the moment of stimulus change was a loud relay report which was audible to *S*s. This sound was constant in pattern and intensity and as such gave no information to *S*s other than that a new stimulus was present. The change from one stimulus to the next was automatically controlled so that the onsets of the visual and auditory portions were essentially simultaneous.

Task and instructions.—The *S*s were told to place each combination of visual and auditory stimuli (called the "total signal") into one of two classes or categories by depressing one of the two levers on the boxes in front of them. It was pointed out that all of the total signals which belonged in the "plus class" (designated by the lever with a plus sign above it) had some aspect or aspects in common and that part of their task was to figure out which aspects of the total signal would allow them to correctly assign a signal to the plus class. The operation of the levers and the corresponding meaning of the information lights were explained to *S*s along with the restriction that they choose one and only one lever to press for each signal.

The objective presented to *S* was to assign as many signals as possible to the correct class. They were told that by finding the solution to the problem (the relevant aspects of the signal) they could greatly increase their accuracy.

During a 1-min. rest in the middle of the session, *S*s were told to relax but to not discuss the problem. Before resuming the problem they were told that the problem was the same and that they should continue with the same objectives.

At the end of the session, each *S* was asked to write down the solution of the problem or any partial solution he had attained.

Stimuli and problems.—The visual stimuli, consisting of linear arrangements of a number of geometric forms, were off-white in color in contrast to the darker background of the paper screen. All visual figures contained a center fixation cross (+). The six dimensions of variation used and both levels of each were: (*a*) Size—large or small; (*b*) Shape—circles or ellipses; (*c*) Number of figures—4 or 3; (*d*) Orientation (linear direction of figures)—horizontal or vertical; (*e*) Location (on screen)—right or left; (*f*) Center spot—below fixation cross or above fixation cross.

The auditory stimuli varied along five dimensions, with two levels of each dimension. Easily discriminated levels for these five dimensions were determined in a prior psychophysical study: (*a*) Pitch of a pure tone—300 or 1800 cps; (*b*) Intensity of the tone—35 or 59 db. above threshold for the 300-cycle tone or 54 or 84 db. above threshold for the 1800-cycle tone; (*c*) Presence or absence of two short "blips" of white noise; (*e*) Steadiness of the pure tone or its regular interruption; (*f*) Earphone to which the signal was delivered— left or right.

When the variation of a dimension was not called for in a given treatment, the dimension was held constant at the level mentioned first in the above descriptions.

The temporal relationships between the noise and the tone interruption are illustrated in Fig. 1. It should be noted that the signal was recurrent, and could have started at any point on a given trial.

Three problems (A, B, and C) were used, selected by randomly pairing three auditory dimensions with three visual dimensions and obtaining three pairs of relevant dimensions. The selection of the particular levels of the relevant dimensions used to define a positive instance of the concept was at random.

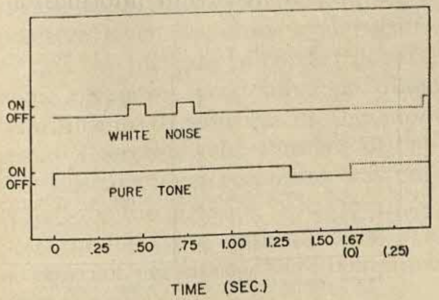


FIG. 1. Temporal relationship between the tone and noise portions of the auditory signal.

TABLE 1

RELEVANT AND IRRELEVANT DIMENSIONS USED IN EACH PROBLEM

Problems	Relevant (Positive Level)		Irrelevant					
	Visual	Auditory	Visual			Auditory		
			1	2	3	1	2	3
A	F (circle)	E (right)	L	O	C	I	W	R
B	S (large)	R (steady)	L	O	C	I	W	E
C	N (four)	P (high)	L	O	C	I	W	E

Note.—The letters in the table refer to the following dimensions: Center spot (C), earphone (E), intensity (I), form (F), location (L), number (N), orientation (O), pitch (P), interruption (R), size (S), and white noise (W).

Table 1 presents the relevant dimensions and the positive instance for each problem as well as the irrelevant dimensions for each. In reading Table 1, the numbered columns under "Irrelevant" indicate the order of adding irrelevant dimensions. Because of a limited selection of auditory stimuli, the third irrelevant auditory dimension in Problem A was different from Problems B and C.

Stimuli were arranged in a series of 96 which was repeated during the second half of the task. The positive instances of the concept, in the same locations for all problems, were placed in the series by random sampling without replacement, as were all other combinations of the stimuli. This allowed for complete balancing of each level for all dimensions as well as for the almost complete balancing of the number of times each level of a given dimension occured with particular levels of all other dimensions. The balancing of levels was done between each mode of presentation (auditory and visual) as well as within each mode. It should be noted that 25% of the stimuli were positive instances of the concept while 75% were negative instances.

RESULTS

Errors.—As the amount of irrelevant visual information was increased

TABLE 2

ANALYSIS OF VARIANCE OF TOTAL ERRORS

Source	*df*	*MS*	*F*
Problems (P)	2	86.391	
Irrel. auditory (A)	(3)	(277.394)	1.37
Linear	1	712.426	3.52
Quadratic	1	.047	
Cubic	1	119.709	
Irrel. visual (V)	(3)	(2772.852)	13.69**
Linear	1	7975.301	39.37**
Quadratic	1	282.755	1.40
Cubic	1	60.501	
P × A	6	139.342	
P × V	6	147.884	
V × A	9	279.079	1.38
P × V × A	18	229.548	1.33
Ss in cells[a]	48	202.589	4.77**
Halves of task (H)	1	9310.255	219.36**
H × P	2	87.099	2.05
H × A	3	8.852	
H × V	3	120.089	2.83*
H × P × A	6	30.175	
H × P × V	6	33.745	
H × V × A	9	38.982	
H × P × V × A	18	14.256	
H × *Ss* in cells[b]	48	42.443	
Total	191		

[a] Error estimate for terms above "*Ss* in cells."
[b] Error estimate for terms containing "Halves."
* $P < .05$.
** $P < .001$.

from 0 to 3 bits, the mean percentage errors increased as follows: 19.3, 21.8, 29.4, and 36.9. An extended trend analysis of variance (Grant, 1956) of the error scores (Table 2) indicated that this increase in errors with increased irrelevant visual information was very significant, and that a significant linear component was the only reliable source of variation. (A Bartlett's test for homogeniety of variance among the conditions yielded a chi square of 58.3 which, with 47 *df* does not force the rejection of the hypothesis of homogeneous variances.) The variations in errors with different amounts of irrelevant auditory information was not significant in this analysis.

The average percentage errors for each *S* was 34.0 for the first half of the task and 18.7 for the second half. This difference was highly significant (Table 2). In addition, there was a significant interaction between the amount of irrelevant visual information and the halves of training. This relationship is shown graphically on the left side of Fig. 2. It can be seen that the decrease in errors from the first to second half of the task became less as the amount of irrelevant visual information in the problem increased.

In addition to the total errors, certain types of errors were of special interest. When the relevant dimension of the visual signal on a given trial was at the positive level, and at the same time the relevant auditory dimension was at the negative level, *S* could have incorrectly selected the positive lever response. This type of error was defined as a Type A error. When the situation was reversed and the relevant visual signal was at the negative level while the relevant auditory signal was positive, the error was called a Type V error. Since the negative levels of each relevant dimension were consistently related to a negative instance of the concept, the Type A errors reflect a tendency of *S* to ignore the relevant auditory information.

These two types of errors were tabulated for each *S* for each half of the task. The *S*s averaged 14.23 Type A errors and only 9.97 Type V errors during the whole task. The differences between the two types of errors were not related to the number of bits of irrelevant information.

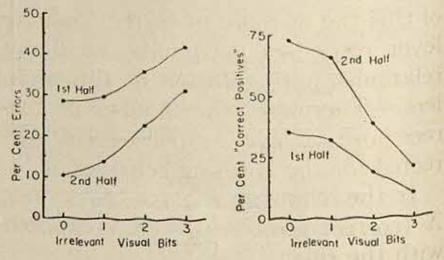


FIG. 2. Performance on first and second halves of the task as a function of the amount of irrelevant visual information.

The difference between the frequencies of the two types of errors was quite significant ($F = 15.96$, 1/80 *df*, $P < .001$) as was the interaction of *S*s by types of errors ($F = 3.39$, 80/80 *df*, $P < .001$). Of interest also is the fact that the interaction of types by halves of the task was not significant ($F = .49$, 1/80 *df*), which would indicate that practice in this task had little if any influence on the predominance of the Type A errors.

Correct positive instances.—Since 75% of the stimuli presented in the task were negative instances, *S*s could substantially reduce their errors by selecting the negative lever response more frequently than the positive response. Moreover, the use of one relevant dimension enabled *S*s to correctly identify half of the instances as negative, because half of the trials contained the negative level of either one of the relevant dimensions. The consistently correct selection of the positive instances, however, required *S*s to take into account both of the relevant dimensions on a given trial. For these reasons, the number of correctly identified positive instances may be a more sensitive measure of the degree of complete solutions attained by *S*s than is a total error or total correct measure. However, the number of correctly identified positive instances, will, on the average, increase as *S* increases the number of positive lever responses and because of this the increase in correct positive lever responses need have no direct relationship to solution of the problem. Therefore, the number of correct positive lever responses was corrected for the guessing behavior of *S*s.

In the following analysis, *S*s' scores, *X* (correct positives), were computed with the rule:

$$X = C^{+} - \frac{C^{-}}{3} \qquad [1]$$

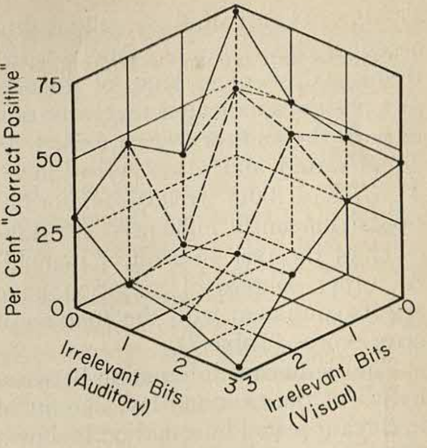


FIG. 3. Relationship between correct positives and amounts of irrelevant auditory and visual information. (Each plotted point represents the mean of six *S*s. The origin of the vertical axis is at the chance level.)

where C^{+} is the number of correct positive lever responses (correctly identified positive instances) and C^{-} is the number of incorrect positive lever responses. The last term of Equation 1, a correction for guessing, was based on the assumption that the average distribution of "guesses" on the positive lever response will be 25% correct and 75% errors. The score, *X*, should then be a good estimate of the number of correct positives above chance performance.

The effects of the amount of visual and auditory irrelevant information on the degree to which *S*s exceeded random behavior in the correct assignment of positive instances is shown in Fig. 3. As the amount of irrelevant visual information was increased from 0 through 3 bits, the mean percentage of correct positives decreased as follows: 53.9, 49.0, 29.0, and 15.3. The corresponding mean values for the auditory dimension were 45.8, 37.5, 34.9, and 29.1. Bartlett's test for homogeniety of variance among the cells gave a chi square of only 28.5 for

47 *df*. This value is well within accepted chance limits. An extended trend analysis of variance of the correct positives indicated that variation in both the auditory ($F = 5.58$, 1/48 *df*, $P < .05$) and visual ($F = 36.98$, 48 *df*, $P < .001$) irrelevant information significantly influenced the difficulty of the task in a linear manner. All other sources of variation gave results consistent with the analysis of error scores (Table 2).

The significant interaction between halves of the task and the amount of irrelevant visual information is shown in the right half of Fig. 2. It can be seen that the increase in correct positives from the first to the second half of the task became progressively smaller as the number of irrelevant visual bits was increased.

Solvers.—On the basis of the written solutions obtained from *S*s at the end of the session, it was quite easy to determine which *S*s had attained adequate verbal solutions to the problems. In addition, the verbal statements clearly indicated partial solutions, mostly the association of the negative level of one relevant dimension with negative instances. Only 36 of the 96 *S*s verbalized the solutions to the problems. The distribution of these solutions was in line with the previous analyses, with the number of solutions decreasing as the number of irrelevant dimensions was increased in either the auditory or visual signals.

Of the 60 *S*s who did not give correct verbal solutions to their problems, 22 correctly identified the relevant visual dimension (at least in its relation to negative instances) while only 4 *S*s did so for the relevant auditory dimension.

DISCUSSION

The obtained linear function relating amount of irrelevant visual information and performance degradation is in line with most of the previous research. An exception is the Archer, Bourne, and Brown (1955) study in which task difficulty increased at a positively accelerated rate as irrelevant information was increased. However, they used stimuli which moved across a screen in a set period of time, whereas all of the other studies with this type of task involved stimuli which were present at least until *S* had responded. It is suggested that the action of a memory factor may account for the relative extreme difficulty in the high complexity conditions of Archer et al. (1955). Under the variety of conditions allowing unlimited observation of stimuli, the difficulty of concept identification is proportional to the amount (in bits) of irrelevant visual information in the stimuli, at least up to 5 bits.

The significant interactions of visual information and halves of the task indicate that the rate of learning decreases as the amount of irrelevant visual information increases. This is consistent with a significant interaction between trials and irrelevant visual information found by Brown and Archer (1956).

Perhaps the most noteworthy result in the present study was the lack of a significant effect (as measured by error scores) of variations in the amount of irrelevant auditory information. The predominance of the Type A errors over the Type V errors indicates that *S*s did not attend to the auditory cues as carefully as to the visual cues. It will also be recalled that this difference did not change significantly between the two halves of the task. It is therefore suggested that *S*s entered the situation with a bias toward using visual information in preference to auditory information and maintained this preference throughout training. This explanation is supported by the verbal reports of a few *S*s who volunteered the information that they had ignored the auditory signal during the session and by the small number of verbal identifications of the relevant auditory dimension.

The measure of percentage of correct positives was intended to provide an indication of the effect of the auditory information with a good deal of the proposed visual bias removed. The significant linear relationship between correct positives and amount of irrelevant auditory information is certainly the type of result expected on the basis of studies with visual stimuli.

The data offer no evidence that there is an interaction between irrelevant visual and auditory information.

Mathematical theories.—Application of statistical learning theories to the concept identification situation have been attempted by Bourne and Pendleton (1958) and by Walker (1958). The formulations of these authors are based on models proposed by Restle (1955, 1957).

The most extensive treatment of the concept situation to date is given by Bourne and Restle (1959). This approach incorporates the earlier applications and makes extensions to more complex situations. Bourne and Restle assume that the relevant cues are perfectly correlated with the reinforcements, i.e., have validity of 1, and they incorporate the cue adaptation model which Restle (1957) has developed.

The present study differs from prior work in that although *S* has only two choices of response he still must consider two stimulus dimensions for solution to the problem. In other words, the average validity of any one dimension is never 1 in the present situation. Restle (1957) has proposed, from empirical results, that the validity of a cue, *V*, be;

$$V = 4\pi^2 - 4\pi + 1, \qquad [2]$$

where π is the proportion of times a given cue is reinforced. Clearly, in the present study, cues from the positive level of a relevant dimension are reinforced on only half of the trials and thus have validity of 0. According to the model, such cues will adapt out of the effective sample. Also the negative level of either relevant dimension can serve only for the identification of negative instances. On this basis, a solution to the problem is not possible! On what basis then are the solvers solving the problem? A possible answer is consideration of cues arising from the combination of the two positive levels from the relevant dimensions (verbal hypotheses?). Such cues have validity of 1 since they are consistently reinforced. The quantitative development of this modification is difficult in view of the additional problems indicated below.

The lack of symmetry in the effect of irrelevant visual information versus irrelevant auditory information does not allow a simple application of the models mentioned above. In these theories, the proportions of effective cues originating from given stimulus dimensions have been taken to be equal (with one exception, color of the stimuli). However, the unequal effectiveness of auditory and visual signals forces some modifications in this assumption.

One additional problem has been created by the present procedure. The so-called irrelevant dimensions in the present study have some validity (assuming Equation 2), for the probability of reinforcement of a given cue from an irrelevant dimension is .25 and not .50 (which would be 0 validity). This problem raises the question of the appropriateness of the term "irrelevant" as used in this paper. Such dimensions are clearly capable of influencing behavior in the manner of the probabilistic matching solutions (Estes, 1950; Restle, 1957) but are not relevant to the *complete* solution possible in the concept task. The implication intended here is that to very sophisticated *S*s cues which allow a fairly high proportion of correct responses may be considered irrelevant when a perfect solution is possible, while to naive *S*s the same cues may be considered as relevant, perhaps as the best possible ones.

SUMMARY

Performance in a concept identification task was studied as a function of the amounts of simultaneously presented auditory and visual irrelevant information.

Groups of 2 *S*s were randomly assigned

to each of 48 conditions. The design involved four levels (0, 1, 2, and 3 bits) of irrelevant visual information, four levels (0, 1, 2, and 3 bits) of irrelevant auditory information, and three problems (1 auditory and 1 visual bit of relevant information in each problem).

The *S*s were instructed to assign each stimulus (consisting of both an auditory and a visual portion) to one of two categories by depressing one of two lever-action switches.

The major results of the study were: (*a*) As the amount of irrelevant visual information was increased, performance decreased in a linear manner. (*b*) The variation in the amount of irrelevant auditory information did not significantly affect the number of errors. (*c*) The number of correctly identified positive instances of the concept (a measure of the degree of complete solution) significantly decreased as a linear function of the amount of irrelevant visual information and as a linear function of the amount of irrelevant auditory information. (*d*) The types of errors made by *S*s gave strong evidence that *S*s responded primarily on the basis of the visual stimuli and tended to ignore the auditory stimuli.

REFERENCES

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

BOURNE, L. E., JR., & PENDELTON, R. B. Concept identification as a function of completeness and probability of information feedback. *J. exp. Psychol.*, 1958, **56**, 413–420.

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

ESTES, W. K. Towards a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.

WALKER, C. M. Concept identification as a function of amounts of relevant and irrelevant information. Unpublished doctoral dissertation, University of Utah, 1958.

(Received July 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 291–294

EYELID CONDITIONING PERFORMANCE AS A FUNCTION OF EMOTION-PRODUCING INSTRUCTIONS[1]

KENNETH W. SPENCE AND HENRY GOLDSTEIN
State University of Iowa

A series of recent studies from the Iowa laboratory has shown that performance level in eyelid conditioning of humans is related to (*a*) UCS intensity (Spence, 1953, 1958; Spence & Taylor, 1951), (*b*) extreme scores on a questionnaire designed to select *S*s differing in the degree to which they manifest emotional symptoms (Spence & Farber, 1953; Spence & Taylor, 1951; Taylor, 1951), (*c*) physiological indices of emotionality of *S*, such as GSR, heart rate, and muscle action potential (Runquist & Ross, 1958, 1959; Runquist & Spence, 1959), and (*d*) shock and/or threat of shock given *S* prior to or during the course of conditioning (Spence, 1958; Spence, Farber, & Taylor, 1954). These findings have been interpreted as reflecting different levels of drive (*D*) which in turn has been conceptualized in terms of different levels of a hypothetical emotional response (r_e) in *S* (Spence, 1958).

As a further test of this theory that the level of the emotional response of *S* determines *D* and hence response strength ($R = fE = H \times D$), an experiment was conducted in which *S*s, after receiving 20 conditioning trials with a weak UCS (air puff to cornea) were informed that the UCS was going to be changed to a very intense value that would be quite unpleasant. On the assumption that such a threat would markedly increase the level of emotional response it was predicted that the conditioning performance of this experimental group would be higher than that of a control group not so threatened. A second experimental group was run which not only was threatened with a stronger puff after Trial 20, but actually received the stronger puff from this point on. Comparison of these two experimental groups provided a measure of the relative effects on performance of the threat alone and the threat plus increased intensity of the puff.

METHOD

Apparatus and method of recording.—The equipment for recording eyelid closure and presenting the UCS was identical with that used in previous studies (cf. Spence & Taylor, 1951). The CS was an increase in brightness of a 6-cm. circular disk from .004 to 2.95 mL. The weak UCS was a 0.33-psi air puff and the strong UCS a puff of 3.0 psi. The UCS was delivered to the right eye and was limited in duration to 50 msec. The CS-UCS interval was 500 msec. with the duration of the CS being 550 msec. A CR was recorded whenever the record showed a deflection of 1 mm. or more in the interval 200–500 msec. following the onset of the CS. Responses with a latency of less than 200 msec., which were infrequent, were classified as original responses, and were not included in the data.

Procedure.—Each *S* was instructed to blink once to a ready signal (the word "ready") and then look at the disk, remaining as relaxed as possible and making no attempt to control the movements of his eyelid. Following the reading of the instructions each *S* received three presentations of the CS alone followed by a single presentation of the UCS. All *S*s were then given 40 conditioning trials. Intertrial intervals of 15, 20, or 25 sec., averaging 20 sec. and arranged by a fixed

---

[1] This research was carried out as a part of a project concerned with the influence of motivation on performance in learning under Contract N9 ONR-93802, Project NR-154-107 between the State University of Iowa and the Office of Naval Research. Acknowledgement is made to David Lucky who ran some of the *S*s.

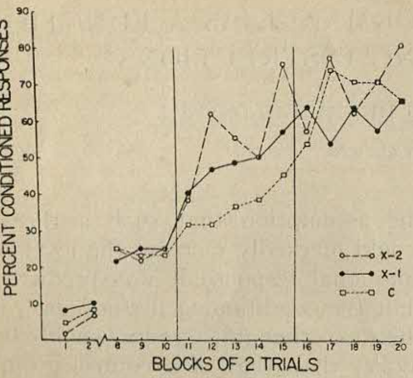


FIG. 1. Mean percentage of CRs in blocks of two trials.

schedule, were used. The interval between the ready signal and CS onset was 2, 3, or 4 sec., randomly varied.

The three groups were differentiated in terms of both the number of the trial after which a threat of the strong UCS was given and the number of the trial on which the strong UCS was actually introduced. All groups received the weak (0.33-psi) puff for the first 20 trials. The control *S*s (Group C) were given 10 further conditioning trials with the weak UCS and then immediately after Trial 30 were threatened with and given the strong UCS for the next 10 trials (31–40). The first experimental group (X-1) received the threat of the strong UCS immediately after Trial 20, but the weak UCS was continued until Trial 30 at which time the strong UCS was introduced and used to the end of conditioning (Trial 40). The second experimental group (X-2) received the threat after Trial 20 with the strong UCS also being introduced immediately (i.e., on Trial 21) and continued until the end of training (Trial 40).

While the critical data in which we were interested was performance on Trials 21–30, the extra 10 trials were run for two reasons. First, so that all *S*s would be shifted to a strong puff and would so report if they discussed the experiment with others. The second reason was that we wished to eliminate so-called "inhibitors," *S*s who showed little or no conditioning even with the strong UCS and Trials 31–40 offered a period in which the drive condition (intensity of UCS) was the same for all *S*s.

The verbal threat consisted, essentially, of telling each *S* that following one of the next 10 presentations of the light (CS), and following every presentation thereafter, he would receive a very strong and unpleasant puff of air (UCS). It was stressed that the first presentation of the stronger puff would be determined by chance (by drawing a number out of a box) and that *S* would have no way of knowing ahead of time on which of the next 10 trials he would actually begin receiving the strong puff. At the conclusion of the session *S* was questioned as to his knowledge of the purpose of the experiment and warned not to discuss the experiment with other members of the class.

Subjects.—The *S*s were 51 men and 39 women students from a course in elementary psychology. At the beginning of the experiment each *S* was assigned to one of three groups in order of appearance at the laboratory. Towards the end of the experiment, *S*s were assigned to the different groups so as to match as closely as possible their average performances on Trials 1–20. Each group then consisted of 30 *S*s, 13 females and 17 males. A number of additional *S*s were discarded for various reasons, 11 because they met the previously established criterion defining a voluntary responder (Spence & Ross, 1959). Nineteen were eliminated as "inhibitors" because they failed to give at least two CRs in the final block of 10 trials, during which all *S*s received the strong UCS. One *S* was eliminated because he gave CRs during the preconditioning trials.

RESULTS AND DISCUSSION

Figure 1 shows the performance curves in terms of the percentage frequency of CRs of the three groups of *S*s for each block of two trials over the course of conditioning. As may be seen, the groups were very closely matched on the first 10 blocks of trials. The differences in performance levels on Blocks 11–15 reflect

TABLE 1

MEAN INCREASE IN PERCENTAGE OF CRS FROM BLOCK OF 10 TRIALS PRECEDING THE INSTRUCTIONS TO BLOCK OF 10 TRIALS FOLLOWING

Group	Mean Percentage Increase	*SD*
C	12.7	3.94
X-1	25.0	4.20
X-2	32.0	3.52

the effects of the different experimental procedures for manipulating level of emotionality. Thus it will be seen that the curve of Group X-2, which was threatened with and actually administered the strong (3-psi) puff on Trials 21–30 jumped up the most. The performance of the group (X-1) only threatened with a strong puff also showed marked acceleration and, as may be seen, its performance was substantially above that of the control group (C) on this block of trials.

Table 1 presents both the mean percentage of CRs from Trials 11–20 to Trials 21–30. Evaluation of the differences between these increases by a simple analysis of variance provided an F of 7.41 ($P < .005$). A comparison of the difference between Group C and Group X-1 yielded a t significant at the .025 level, while that between Group C and Group X-2 was significant at the .001 level. The difference between the two experimental groups was not significant ($P = .22$).

These findings clearly support the prediction based on our emotional response theory of drive (D) that the level of conditioning performance of the experimental group (X-1) threatened with a strong, unpleasant puff would be significantly higher than the control group (C). In connection with this finding it is interesting to recall the results of Taylor (1951). In a study primarily concerned with the relation between conditioning performance and the level of emotionality of *S*s as determined by a questionnaire-type scale, Taylor also attempted to manipulate experimentally the emotionality level of her *S*s after 20 conditioning trials. Half of each type of her *S*s was instructed that the intensity of the UCS would be increased on subsequent trials and half told that the intensity would be decreased. Actually the intensity of the UCS remained unchanged.

TABLE 2

COMPARISON OF RESULTS OF PRESENT STUDY WITH THOSE OF TAYLOR (1951)

Study	Group	Mean Percentage Increase of CRs	
		Trials 21–25 vs. 11–20	Trials 26–30 vs. 11–20
Present	C	8.7	16.7
Present	X-1	21.3	28.7
Taylor	Plus	17.0	11.4

Taylor found that the mean increase in percentage of CRs for Trials 21–30 over Trials 11–20 was 14.2 for her two groups (high and low emotionality *S*s combined) that were threatened with an increase in puff intensity. This is to be compared with 25% in the case of our threatened group. Suspecting that the effects of her threat instructions had only a temporary effect, Taylor further analyzed her data and found corroborative evidence in that the increase over Trials 11–20 for this threatened group was, as shown in Table 2, 17.0% for Trials 21–25, whereas in Trials 26–30 there was a drop in performance to a level only 11.4% above that of Trials 11–20.

Anticipating the development of this scepticism of a change in puff intensity on the part of *S*, we attempted to prevent its occurrence by framing our instructions so that *S* would continue to expect the change even when it was observed not to occur over the 10 trials. That our procedure was successful in this respect is revealed by the fact that our threatened group (X-1) continued to show an increase in the second block of five trials. Thus, as revealed in Table 2, Group X-1 showed a larger increase over Trials 11–20 on Trials 26–30 (28.7%) than on Trials 21–25 (21.3%). As may be seen by comparing these latter values with the comparable ones for the control group (C), the amount by which the performance of the threatened group was elevated over that of the control group is approximately the same in both blocks of trials (12.6 vs. 12.0).

SUMMARY

This experiment was conducted to ascertain the effects of introducing emotion-elevating instructions during the course of an eyelid conditioning experiment. One group of *S*s that had received 20 conditioning trials with a weak UCS (.33-psi puff) was informed that the UCS would be increased to a very unpleasant intensity at some time during the following 10 trials. On the assumption that this threat would materially increase the level of emotional response (r_e) and hence the level of *D* of *S*, it was predicted that the performance of this group over this block of trials would be significantly higher than that of a control group not given such a threat. This prediction was confirmed. A third group was run in which the UCS was actually changed to a rather strong puff (3.0 psi) after 20 trials with the weak puff. The performance level of this group on the test trials (21–30) was slightly, but not significantly, above that of the threatened group. Thus the drive-increasing effect of the instructions was slightly less than that produced by increasing the UCS from a puff of .33 psi to one of 3.0 psi.

REFERENCES

RUNQUIST, W. N., & ROSS, L. E. A preliminary investigation of the relationship between GSR, heart rate change, and eyelid conditioning. *Proc. Iowa Acad. Sci.*, 1958, **65**, 398–404.

RUNQUIST, W. N., & ROSS, L. E. The relation between physiological measures of emotionality and performance in eyelid conditioning. *J. exp. Psychol.*, 1959, **57**, 329–332.

RUNQUIST, W. N., & SPENCE, K. W. Performance in eyelid conditioning related to changes in muscular tension and physiological measures of emotionality. *J. exp. Psychol.*, 1959, **58**, 417–422.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

SPENCE, K. W. A theory of emotionally based drive and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

SPENCE, K. W., FARBER, I. E., & TAYLOR, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

SPENCE, K. W., & TAYLOR, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TAYLOR, J. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

(Received July 16, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 295–301

PROACTIVE INHIBITION OF CONNECTED DISCOURSE[1]

NORMAN J. SLAMECKA

University of Vermont

The phenomenon of proactive inhibition (PI) has never been reported for the rote retention of prose materials (Slamecka & Ceraso, 1960). The main purpose of this paper is to determine whether passages of connected discourse are in fact subject to such PI, and also to examine the effects of two variables upon the magnitude of such interference. One of these variables is the amount of prior learning (PL), expressed in terms of the number of interfering passages learned. With paired associates, Underwood (1945) found that PI increased directly with increases in the number of prior lists learned. The other variable under investigation is the interval between the learning and test for retention of Passage 2. According to logical expectations, PI should increase with time (up to some undetermined point), and the available evidence suggests that this is probably the case, at least up to 48 hr. (Slamecka & Ceraso, 1960). Since our previous studies have indicated the generalizability of findings based upon nonsense materials to prose passages (Slamecka 1960a, 1960b), we made the following hypotheses about the present experiment: (*a*) the rote retention of passages of connected discourse is subject to significant PI, (*b*) PI varies directly with the number of previous passages learned, and (*c*) PI increases with increasing retention intervals.

In addition, this study is also concerned with another aspect of knowledge about retention, strikingly pointed out by Underwood (1957). By summarizing retention data from a number of studies, he showed that the percentage recalled after 24 hr. decreased directly as a function of *S*'s stage of practice, or the number of lists previously mastered. Thus, PI effects (through the use of highly practiced *S*s) have served to depress the amount recalled below that actually to be expected with naive *S*s, and the commonly published retention curves found in textbooks are confounded with such effects. Underwood (1957) concludes that "we must start all over and derive a retention curve over time when the subjects have learned no previous materials in the laboratory" (p. 56). To our knowledge there is in the literature no such curve based upon connected discourse. Underwood concludes that the loss in retention of verbal materials is only 25% over 24 hr. with naive *S*s, and that even this value is inflated by the failure to take into account the postcriterional drop in performance at the conclusion of learning. The present study shall provide data for four retention curves covering a 24-hr. period, one of which is a curve produced by naive *S*s, allowing an empirical comparison of Underwood's findings to connected discourse.

METHOD

Subjects and materials.—Thirty-six students in general psychology at the University of Vermont served as *S*s as part of the requirements of the course. Every *S* participated in four individual sessions, during each of which he memorized a prose passage and was tested for subsequent retention after a predetermined interval. The materials were four fairly difficult sentences, each exactly 20 words in length, all rather similar in subject matter. The passages used were:

---

[1] This research was sponsored by National Science Foundation Grant G-6192. The author is indebted to John Corson and to William Adams, who assisted in the collection and analysis of the data.

1. "Any generalized statement of relationship between variables requires specification of some reference class or universe from which the material comes."
2. "According to recent considerations it appears preferable to select preliminary assumptions from the realm of objectively verifiable and observable phenomena."
3. "Traditional rationalism cannot negate the assertion that its theory concerning primary evidence rests upon inadequate analysis of fundamental scientific methodology."
4. "Forthright admission of the contingent character of our most soundly established beliefs distinguishes contemporary empiricists from dogmatic and positive ancients."

The words of the passages were presented on a memory drum at a 3-sec. presentation rate and a 6-sec. intertrial interval. Learning was by serial anticipation.

Design and procedure.—A mixed factorial design (Lindquist, 1953, Type II) with four levels of each variable was used. One variable was the number of prior passages learned (0, 1, 2, and 3), and the other was the retention interval (15, 30, and 60 min., and 24 hr.). Of the 16 combinations of these levels, each *S* had four, in such a manner that he had all levels of each variable but no repetitions of any level. The 36 *S*s provided 9 replications of the 16-cell matrix. Conditions were arranged so that passages were counterbalanced with respect to the two main variables. Sessions, of course, were counterbalanced only with respect to retention intervals and passages but not with amount of PL, since the session sequence determined the PL level, i.e., Sessions 1, 2, 3, and 4 were always associated, respectively, with 0, 1, 2, and 3 previous passages.

Each experimental session began with one warm-up trial of number guessing on the memory drum. Next, instructions for serial anticipation were given and then the designated passage was presented. The criterion for original learning (OL) was one errorless trial plus one additional trial.[2] The designated retention interval was passed outside of the laboratory. For the 15-, 30-, and 60-min. intervals *E* engaged *S* in casual conversation, and for the 24-hr. interval *S* was dismissed until the following day. Upon return to the laboratory, *S* was given one trial of number guessing for warm-up, followed by relearning of the last passage. The criterion for relearning was one errorless trial, but all *S*s were given a minimum of five trials, regardless of whether they may have relearned in fewer than that. As soon as relearning of one passage was complete, *S* was given the next passage to learn, and the above procedure was repeated until each *S* had successively acquired and relearned all four passages.

RESULTS AND DISCUSSION

The acquisition data were evaluated by way of three criteria. Mean total correct anticipations per *S* are presented in the upper half of Table 1, and the results of the statistical analysis are found in the left half of Table 2. It is seen that amount of PL had a significant facilitating effect in that the better practiced *S*s needed fewer reinforcements to learn the passages.

Acquisition performance was also measured by total trials during learning, and by performance level on the last trial. Mean total trials as a function of increasing PL levels were, 14.2, 11.8, 10.5, and 10.1. These are significant beyond the .001 level ($F = 22.52$, $df = 3/96$). Mean numbers correct on the last OL trial (the trial following the errorless criterion trial) with increasing amount of PL were 18.61, 19.11, 19.39, and 19.42. The expected postcriterional drop did occur, and although the differences were slight the means were significant beyond the .05 level ($F = 3.44$, $df = 3/96$). All three analyses are consistent in showing a significant

---

[2] This would provide information on the postcriterional drop for prose materials. Hilgard (1951) says: "Any criterion that depends upon the first reaching of some high level of performance capitalizes on chance fluctuations in the direction of good scores. The mean of scores just after the criterion will tend to fall below the criterion trial" (p. 554).

TABLE 1

Mean Total Correct Anticipations during Acquisition and Mean Recall Scores

Prior Learning (Lists)	Retention Intervals							
	15 min.		30 min.		60 min.		24 hr.	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
	Acquisition							
0	214.3	93.7	190.6	50.4	179.1	39.7	160.8	55.8
1	125.7	28.2	134.3	45.2	153.7	56.2	189.8	34.8
2	147.3	39.6	121.7	40.5	170.3	76.8	138.2	50.4
3	124.0	25.9	153.6	30.9	132.6	35.0	131.9	44.7
	Recall							
0	18.1	1.2	17.9	1.9	17.8	1.9	15.2	3.2
1	18.1	2.4	16.8	1.5	15.4	2.7	13.3	3.1
2	15.2	3.3	16.1	3.2	13.7	3.6	11.7	3.5
3	13.7	3.6	13.3	2.9	13.2	3.4	9.3	4.9

practice effect for acquisition of prose materials, and are in good accord with similar facilitating effects commonly reported for nonsense items. None of the groups differed on original learning as a function of the subsequent retention intervals.

The main analysis of retention was through the recall (first relearning trial) scores. Recall data are presented in the lower half of Table 1, and the statistical analysis summary is in the right half of Table 2. The effects of amount of PL and retention intervals were significant. Appropriate *t* tests indicated that all adjacent means differed from each other at or beyond the .025 level for the PL variable, and with retention interval only the 24-hr. group differed significantly from all others ($P < .01$). Because of the possibility of heterogeneity of variance in these data, supplementary analyses were performed using the distribution-free Friedman analysis of variance of

TABLE 2

Analyses of Variance for Total Correct Anticipations during Acquisition, and for Mean Recall Scores

Source	Acquisition			Recall		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *S*s	35	6893.8	.99	35	12.1	.94
PL × Int. (b)	3	6136.1	.88	3	3.9	.31
Error (b)	32	6964.8		32	12.8	
Within *S*s	108	2322.9	1.35	108	18.1	1.55*
PL	3	25572.7	14.90**	3	160.8	13.75**
Interval	3	489.1	.28	3	113.8	9.74**
PL × Int. (w)	6	1313.6	.76	6	1.8	.16
Error (w)	96	1716.8		96	11.7	

* $P < .05$.
** $P < .01$.

TABLE 3

Relationship between Acquisition Rates and Recall

Prior Learning (Lists)	Total Correct Anticipations per S in Acquisition							
	50–99		100–149		150–199		200-more	
	N	Mean	N	Mean	N	Mean	N	Mean
0	0	—	12	17.3	12	17.2	12	17.2
1	7	16.9	11	16.6	11	15.7	7	14.1
2	10	13.0	11	16.1	10	14.3	5	12.0
3	9	11.8	13	12.3	12	13.3	2	10.0

ranks. A χ_r^2 of 34.3 for PL ($P < .001$, $df = 3$), and of 20.4 for retention interval ($P < .001, df = 3$), confirmed the effectiveness of both variables.

At this point in the analysis the problem of the role of practice effects upon retention had to be examined. It could be argued that the differences at recall were not true PI effects, but rather were the result of differential acquisition rates. That is, *S*s who needed more total trials or correct anticipations during acquisition of a passage should recall it better than those who took fewer trials, simply because of their more intensive training upon it. Two approaches were used to investigate this possibility. First, within each PL level the correlation between total correct anticipations and recall was calculated. If the degree of specific training on a passage did indeed determine its retention, there would be a significant positive correlation, and the recall scores would have to be adjusted accordingly. The obtained *r*'s with increasing PL levels were respectively, .09, −.16, −.19, and .06 (all at $df = 35$). None of these approached significance, thereby indicating no need for adjustments and no evidence for a within-PL groups effect.

The second method, originally used by Underwood (1945), involved an across-PL groups comparison of recall, with each PL group subdivided into categories of total correct anticipations (or total reinforcements). The relevant data are found in Table 3. If it were the case that recall level was determined by acquisition rate, then the recall means would rise from left to right across the table, and also would remain constant vertically, across PL levels. Inspection of the table shows that this is not so, but rather that within each total correct anticipation category the recall means decrease steadily with increasing PL levels as would be expected of true PI effects.

These data are also consistent with Underwood's (1954) conclusion that fast and slow learners do not differ in retention. At first glance it may appear, paradoxically, that amount of reinforcement during acquisition has no effect upon retention, but this is only the result of our sorting out *S*s on the basis of learning rate. Studies specifically designed to demonstrate the effects of various degrees of learning upon retention (Briggs, 1957; Slamecka, 1960a) usually employ a fixed trials acquisition procedure wherein each treatment group is presumably heterogeneous with respect to fast and slow learners.

From the two preceding analyses we conclude that the recall data prove the presence of genuine PI, which confirms the main hypothesis of this study, and that the degree of PI increases with increasing numbers of previously learned passages, thus confirming the second hypothesis. Expressed in terms of the formula, (Rest-Work/Rest) × 100, relative PI as a function of increasing amounts of PL was found to be 8.0%, 18.1%, and 28.6%. Relative PI as a function of increasing retention intervals was 13.5%, 13.9%, 20.6%, and 24.8%, thus confirming the third hypothesis that PI increases with time, at least up to 24 hr.

Analysis of total trials to relearn showed the usual transitoriness of PI in that there was no significant effect of amount of PL ($F = .85$, $df = 3/96$),

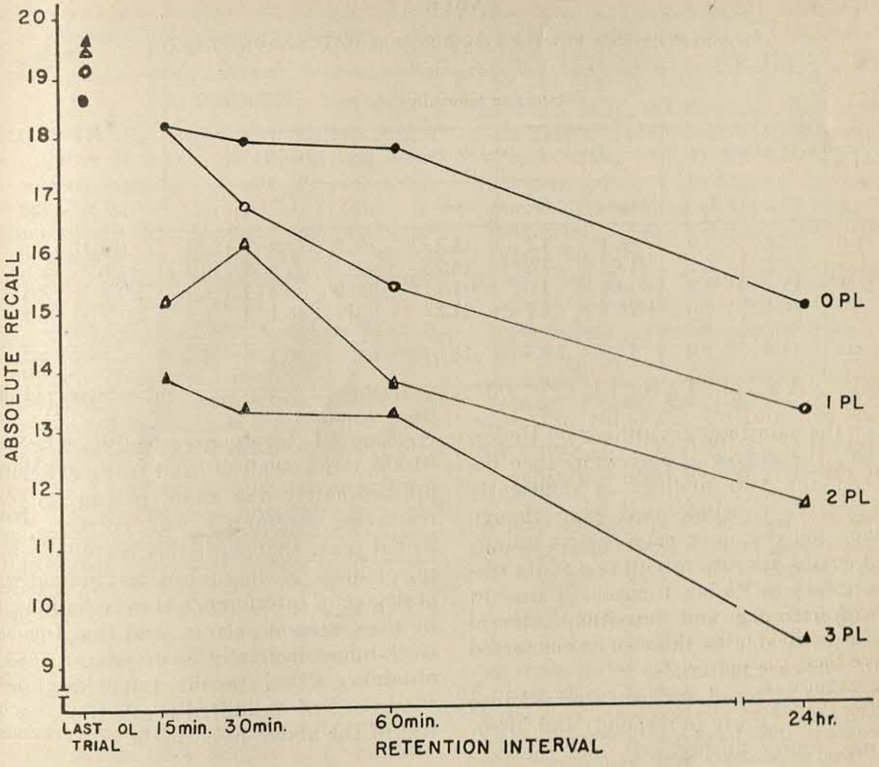


FIG. 1. Absolute amount recalled as a function of amount of prior learning (PL) and retention interval.

but retention interval was significant beyond the .01 level ($F = 5.87$, $df = 3/96$).

Retention curves in terms of absolute amount recalled, together with the terminal OL trial performance of each group are found in Fig. 1. The uppermost line is the curve for naive *S*s. In spite of the fact that recall on the postcriterional trial was better the greater the amount of PL, and that all of the long-duration recalls are inflated somewhat by the learning on that last trial, the orderly decremental effects of increasing PI are clearly evident at each retention interval. These results strongly suggest that the classical Ebbinghaus type of curve, with its steep initial drop and considerable overall decrement (most closely approximated by our 3-PL group) is a function of cumulative proactive interference attributable to practiced *S*s.

Underwood (1957) estimated that naive *S*s actually forgot only 15% over 24 hr., with the assumption of a postcriterional drop of about 10%. Our data show that the mean postcriterional drop for naive *S*s was 1.39, or 7%, which is in fair accord with Underwood's figure. Then, using the values of 18.61 (mean performance of naive *S*s on final learning trial) and 15.22 (mean recall of naive *S*s after 24 hr.) we obtain a retention loss of 18.2%, which is a slight underestimate because of the unmeasured learning during the postcriterial trial. Considering the variability resulting

TABLE 4

PERCENTAGES OF OMISSION AND INTRUSION ERRORS AT RECALL

Prior (Lists)	Retention Interval								Mean	
	15 min.		30 min.		60 min.		24 hr.			
	Om.	Int.	Om.	Int.	Om.	Int.	Om.	Int.	Om.	Int.
0	5.6	3.9	6.1	4.4	8.3	2.8	17.8	6.1	9.4	4.3
1	7.2	2.2	13.3	2.8	16.7	6.1	22.8	10.6	15.0	5.4
2	19.4	4.4	14.4	5.0	21.7	10.0	35.6	6.1	22.8	6.4
3	26.1	5.6	26.7	6.7	27.8	6.1	46.7	6.7	31.8	6.3
Mean	14.6	4.0	15.1	4.7	18.6	6.2	30.7	7.4		

from the many sources utilized by Underwood, the degree of correspondence between the two findings is sufficiently high. We conclude that even though connected discourse may possess unique properties, its rote retention and its susceptibility to PI as a function of amount of interference and retention interval are comparable to those of unconnected and nonsense materials.

Errors at recall were categorized into omissions (failure to respond) and intrusions (source unspecified), and the percentage occurrence of each is given in Table 4. Examination of the table shows that intrusions played a very minor role throughout, and that they increased only slightly with increases in interference levels and retention intervals. As in previous studies (Slamecka, 1960a, 1960b), omission errors accounted for the bulk of the changes at recall, for they increased markedly with increases in PL and with lengthening retention intervals. This again points to the hypothesis that with serial lists of connected discourse material, the main effect of interfering lists is to increase the reaction latencies of the correct resonses, and with the use of limited response times (3 sec. in this study), the obtained omission pattern would follow. Further support for the contention that interference affected mainly the latencies but not the actual retention of the items in given by calculating the percentage of all overt responses given at recall, that were correct. These accuracy scores, in order of increasing PL levels, are: 95.2%, 93.6%, 91.7%, and 90.9%, and suggest that approximately the same percentage of responses attempted was correct. It would seem that with this material, the importance of intrusions as indicators of degree of interference is overshadowed by the omission pattern, and that future work might fruitfully be directed toward obtaining actual response latencies under conditions of unlimited recall time, as a test of the above hypothesis.

SUMMARY

This study tested the susceptibility of the rote retention of prose passages to proactive inhibition (PI) of recall, as a function of amount of prior learning (PL) and length of retention interval. Thirty-six *S*s each learned and recalled four 20-word passages of connected discourse in succession, by the method of serial anticipation. There were four levels of amount of PL (0, 1, 2, and 3 passages), and retention intervals were 15, 30, and 60 min., and 24 hr. The following results were obtained: (*a*) Significant practice effects occurred during acquisition. (*b*) Significant PI of recall was obtained. (*c*) PI increased directly with the amount of PL. (*d*) PI increased directly with the retention interval. (*e*) After 24 hr. naive *S*s forgot about 18%, thus supporting Underwood's findings about this point (*f*) Omission errors at recall were more predictive of the retention pattern than were intrusions.

It was concluded that whatever unique properties may be possessed by connected discourse, it also shows the same susceptibility to PI effects as does nonsense material. A hypothesis that interference of this kind

serves mainly to increase the latency but not the accuracy of correct responses, was discussed.

REFERENCES

BRIGGS, G. E. Retroactive inhibition as a function of degree of original and interpolated learning. *J. exp. Psychol.*, 1957, **53**, 60–67.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of practice level. *J. exp. Psychol.*, 1960, **59**, 104–108. (a)

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of similarity of topic. *J. exp. Psychol.*, 1960, **60**, 245–249. (b)

SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 273).

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

(Received July 18, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 302–309

RETENTION OF INFORMATION UNDER CONDITIONS APPROACHING A STEADY STATE[1]

ROGER N. SHEPARD AND MARTHA TEGHTSOONIAN
Bell Telephone Laboratories *Eastern Pennsylvania Psychiatric Institute*

In experiments on human memory, *S*s are typically presented with an isolated block of information and then tested for retention of that information. The digit span test is a familiar example. In experiments of this kind *S*s have recourse to various devices of covert regrouping, recoding, and rehearsal of the information. Furthermore, interference resulting from prior exposure to similar information is usually minimized. The results of such experiments may not, therefore, be generalizable to situations that confront *S*s with a continuing sequence of items and require them, throughout the sequence, to retain as much as possible of the most recently presented information. Situations of this type are nevertheless quite common (as, for example, in playing a hand of bridge) and in some cases they are of considerable practical importance (as in the continuous monitoring of a complex display). They also have a wide theoretical bearing; for in such varied activities as inductive inference, deductive reasoning, problem solving, and verbal communication *S*s' responses at any one time must be selectively modulated by stimuli variously strung out into the recent past. Cahill and Hovland (1960) have demonstrated the crucial role that this kind of memory plays during the acquisition of concepts.

Recent efforts to measure retention during continuous tasks have employed recall procedures (e.g., Mackworth, 1959; Pollack, Johnson, & Knaff, 1959). In the present experiment, however, a recognition procedure is selected instead. Unlike the usual recall procedure, the recognition procedure enables *E* to single out, for test, the one item presented at any specified earlier point in the sequence and so to vary at will the number of items intervening between the initial and test presentation of an item.[2] By testing after different numbers of intervening items, then, the course of forgetting of single items during the presentation of subsequent items can be inferred. A curve of forgetting can thus be obtained under conditions that depart from those usually imposed in that the interval between the original and test presentations of an item, far from being filled with an entirely different activity, is filled simply with further presentations of exactly the same kind. In this way an extreme can be approached in which the possibility of rehearsal is minimized while the interference of preceding material is maximized.

---

[1] The present experiment was conducted at Harvard University under Contract AF33(038)-14343 with the Operational Applications Laboratory of the Air Force Cambridge Research Center. Reproduction for any purposes of the United States Government is therefore permitted.

This work was reported in April 1959 at the Annual Meeting of the Eastern Psychological Association. The authors are indebted to George A. Miller who furthered this work by encouragement and specific suggestions.

[2] Since the present report was originally prepared several other studies have appeared that use various ingenious methods to achieve a similar result with recall procedures (Anderson, 1960; Lloyd, Reid, & Feallock, 1960; Yntema & Mueser, 1960). The last two of these three studies used continuous tasks that, except for the utilization of a recall rather than a recognition test, were quite similar to the task used here. Indeed, Yntema and Mueser present curves of forgetting that seem to be the analogues, for the recall process, of the curve illustrated in the present Fig. 1.

In addition to making an empirical determination of the curve of forgetting under such conditions, there are two other aims: First, to show that a rational account can be given for these data in terms of a stochastic model originally developed to account for the shape of the gradient of stimulus generalization. And, second, to estimate the total amount of information (about all preceding stimuli of the sequence) that is carried along when steady state is reached; i.e., when the amount of information gained from each successive item just compensates for the amount forgotten from all preceding items.

METHOD

Subjects

The *S*s were 63 students at Harvard University. Of these, the 15 serving in the "variable delay condition" were graduate students in psychology, and the 48 serving in the "fixed delay condition" were summer session students who were paid for their participation.

Procedure

At the beginning of each experimental session *S* was given a large deck of 3 × 5-in. cards each bearing a single three-digit number. Following a brief description of the manner of construction of the deck, *S* was given a specially prepared record sheet and instructed to examine in turn the number on each card and to place a mark on the record sheet either in a column labeled "old" or in a column labeled "new" according as he did or did not remember having seen that number on an earlier card of the deck. He was allowed to proceed through the deck at his own rate but was not permitted to look back to any earlier card. No feedback as to the correctness of his responses was provided.

The three-digit numbers included in each deck were randomly selected from the total population from 000 through 999. However, with the exception of a few nonduplicated numbers toward the end of the deck, every number selected for a given deck occurred exactly twice in that deck. The second occurrences of the numbers were inserted in such a way that (after the first few cards) each successive card bore either an old or a new number with equal probabilities and according to a random sequence. The insertion of a few nonduplicated numbers toward the end was designed to prevent the probability of finding a new number from falling below .5 for the last cards.

The dependent variable was the probability of classification of the number on a given card as "old." It thus became the probability of a correct response, if the number on the given card had also appeared on an earlier card, and the probability of an incorrect response (or "false alarm"), if the number had not. There were two independent variables. The secondary independent variable, *T*, was simply the total number of other cards preceding a given card in the same deck. The primary independent variable, *d* ("delay"), applied strictly only when the number appearing on a given card had appeared on an earlier card. In that case it was defined as the number of other cards intervening between the given card and the earlier card bearing the same number. (Since the rate of presentation was not controlled, delay, as defined here, had only a rough relation to elapsed time. No attempt was made to separate experimentally the effect of the number of intervening presentations per se from the effect of pure temporal delay.) The number of intervening presentations was manipulated differently under the two following experimental conditions:

Variable delay condition.—Four decks were prepared for this condition. In each of these decks the delays varied from 0 to about 60 intervening cards for different three-digit numbers within the same deck. Insofar as possible a roughly rectangular distribution of delays was maintained throughout the decks. (Of course the delays were limited from above for old numbers occurring near the beginning of a deck.) Two of the decks, both containing 100 cards, were each given to 6 *S*s. The other two decks, both containing 200 cards, were each given to 10 *S*s. Although some of these *S*s served more than once, their exposure to different decks was always separated by 2 or more days. The results, moreover, failed to provide evidence of any interdeck transfer.

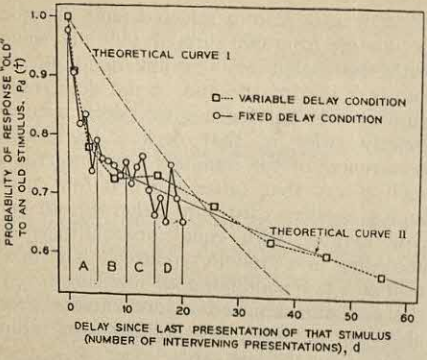


FIG. 1. Probability of classification of an old three-digit number as "old" as a function of the number of intervening three-digit numbers. (Each of the four decks of the variable delay condition was weighted, in calculating the average points, by the number of *S*s times the number of cards for that deck. Of the 21 points for the fixed delay condition, the 18 falling inside one of the four ranges of delays—A, B, C, or D—are each averaged from the two decks covering that range and the 3 points at the intersection of two adjacent ranges are each averaged from the four decks overlapping at that delay. The smooth curves in this and subsequent figures are based upon theoretical considerations presented in the theoretical analysis and discussion section.)

Fixed delay condition.—Eight decks were prepared for this condition. In each of these decks the delays never varied more than $2\frac{1}{2}$ cards from the mean delay for the deck. Different decks realized different mean delays with approximately rectangular distributions for the following ranges: from 0 to 5 for Deck A_1, from 5 to 10 for B_1, from 10 to 15 for C_1, and from 15 to 20 for D_1. The number of cards in the decks varied from 200 to 217, but the number of different three-digit numbers that occurred twice in a deck was held constant at 100. Each of 24 *S*s went through all four of these decks according to a different one of the 24 possible orders. This complete counterbalancing provided for a separation of the effects of interdeck transfer and mean delay. The administration of successive decks to the same *S* was always separated by at least 1 day and in most cases by 2 or more days.

In a replication, 24 new *S*s went through Decks A_2, B_2, C_2, and D_2 according to the same counterbalanced design. These four new decks were prepared in the same way as the first four decks. The independent construction of new decks for the replication, however, insured a more adequate sampling of possible sequences of new and old numbers. In the fixed delay condition the probabilities of new or old numbers were necessarily not completely independent for neighboring cards. The restriction of the delays to a range of five cards for any one deck entailed small sequential dependencies. However, since the variable delay condition—for which there were no sequential constraints—yielded the same results, it seems unlikely that *S*s made any appreciable use of these constraints.

RESULTS

Effect of experience with previous decks.—The performance of *S*s on a given deck did not appear to depend upon the number of other decks to which they had already been exposed. In the fixed delay condition, the percentage of the total classifications (as "old" or "new") that were correctly made by *S*s ($N = 48$) in the four sessions, in the order of their occurrence, were 76, 76, 76, and 77% correct. Since practice evidently had little effect, the data from all four sessions were pooled for subsequent analyses.

Effect of delay, d.—Figure 1 summarizes the effect of delay upon the accuracy of classification of an old three-digit number as "old." The two methods of varying delay yielded essentially indistinguishable results. The probability that a number was classified as "old" dropped rapidly from almost 1.00 (when that number was just seen on the immediately preceding card) to about .75 (when that number had not been seen for 6 or 7 cards). From there on this probability dropped more slowly to about .57 (when the number had not been seen for as many as 50 or 60 cards). However, even for the longest delays imposed, this probability was well above the probability that a *new* number was classified as "old" (the "false alarm" probability), which averaged only about 0.22.

Effect of number of cards already examined in the same deck, T.—Figure 1 might at first be thought to represent the course of forgetting of a single three-digit number under steady state conditions. However, comparisons between the performances of *S*s at different stages of their progress through a deck revealed that a strictly steady state was not completely achieved. In particular an analysis of the data from the fixed delay condition, presented in Fig. 2, clearly showed that the probability of classification of a new number as "old" (the probability of "false alarm"), rather than remaining constant throughout a deck, was still increasing after as many as 200 presentations. However, since the rate of increase of this probability after the first 40 cards was less than .0007 per card, the conditions of the experiment are regarded as "approaching a steady state."

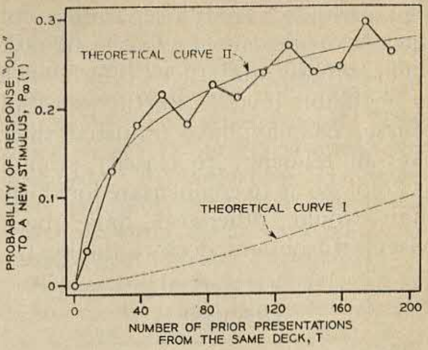


FIG. 2. Probability of classification of a new three-digit number as "old" as a function of the total number of three-digit numbers already seen in the same deck. (The points are averaged from all eight decks of the fixed delay condition.)

The fact that the probability of classifying a new number as "old" increased with the number of other three-digit numbers already examined raises, again, the question of the effect of experience with previous decks. If the trend illustrated in Fig. 2 had continued from one deck to the next (as well as within the same deck), the probability of classifying a new number as "old" should have increased as a function of the number of decks to which *S* had already been exposed. However the results for the fixed delay condition, which are summarized in Table 1, suggest that if the probability of the response "old" changed at all over the four successive decks it probably decreased rather than increased. (This trend, however, is not statistically significant.)

TABLE 1

EFFECT OF MEAN DELAY AND ORDER OF PRESENTATION OF A DECK UPON THE TWO TYPES OF ERROR PROBABILITIES FOR CARDS OF THAT DECK

Order of Presentation of Deck	Probability that New Number is Called "Old"					Probability that Old Number is Called "New"					Mean Probability for Errors of Both Types
	Mean Delay for Deck				Mean Probability	Mean Delay for Deck				Mean Probability	
	A 2.5	B 7.5	C 12.5	D 17.5		A 2.5	B 7.5	C 12.5	D 17.5		
1	.18	.21	.27	.29	.25	.15	.22	.24	.30	.23	.24
2	.17	.26	.24	.30	.24	.17	.26	.25	.28	.24	.24
3	.14	.22	.26	.25	.22	.14	.27	.27	.35	.26	.24
4	.17	.18	.22	.22	.20	.17	.25	.30	.33	.27	.23
Mean	.17	.22	.25	.27	.23	.16	.25	.27	.32	.25	.24

The median 2-day separation of successive sessions may have sufficed, then, to reduce interdeck interference to negligible levels. Furthermore, of course, *S*s may have adjusted their over-all tendency to classify stimuli as "old" so as to compensate for shifts that would otherwise have been observed from one deck to the next.

DISCUSSION

The objectives of this section are to explore some implications of the data presented in Fig. 1 and 2 both for theories of forgetting and for estimates of the information capacity of human memory.

Implications for trace-decay models.—A stochastic model originally developed to account for the shape of the gradient of generalization (Shepard, 1958) has recently been extended to account for phenomena of recognition memory as well (Shepard, 1961). This extended model (referred to here as "the trace model") differs in two respects from previous models of the memory process. First, most previous models predict that the curve of forgetting is a simple exponential decay function with, perhaps, a positive asymptote (Bourne & Restle, 1959; Bower, 1959; Estes, 1955; Hull, Hovland, Ross, Hall, Perkins, & Fitch, 1940, pp. 43, 65; London, 1950; von Foerster, 1950; Witte, 1959; and, in a somewhat different form, Björkman, 1958, pp. 53–55). In contrast to these, the present trace model predicts a simple exponential only under the very special condition that all the presented stimuli are equally similar to each other. If the stimuli vary substantially in similarity, the predicted curve of forgetting departs from the best fitting simple exponential function in that it drops more rapidly at first and then more slowly than that function. Interestingly, then, many previously obtained empirical curves do appear to deviate from a simple exponential in just this way (see, for example, Hanawalt, 1937, p. 25; Jenkins & Dallenbach, 1924, p. 609; Luh, 1922, p. 22; Strong, 1913, p. 350; Williams, 1926, p. 373). That this is also the case for the present data is shown in Fig. 1. There, the trace model was fitted on the assumption that all three-digit numbers are equally similar. The resulting exponential decay function (Theoretical Curve I) clearly departs from the empirical points.

The second distinguishing feature of the present trace model is that it provides an account for the way in which the probability of "false alarm" increases as more and more stimuli are presented (as shown in Fig. 2). This follows from the assumption of the model that forgetting results not only from the fading or simple attrition of traces but also from the spreading out or "diffusion" of traces to other similar stimuli. The specific results of the attempt to fit this model to the present data will now be described.

The probability that a stimulus will be classified as "old" is denoted by $P_d(T)$. Figure 1, then, plots the mean values of $P_d(T)$ (i.e., $P_d(T)$ averaged over all positions in the deck from $T = 0$ through, say, $T = 200$) against delay, d. Figure 2 plots the probability of false alarm, which (according to the model) is given by $P_\infty(T)$, against the number of preceding presentations of stimuli, T. In order to fit the model, a value had to be determined for the response bias, B. Since *S*s were not explicitly instructed to adopt either a stringent or lenient criterion in deciding whether each three-digit number was old, the initial expectation was that there would be no over-all bias. However, when B was set to unity (the null bias), even the most favorable setting of the remaining parameters failed to eliminate a clear departure from the data of Fig. 2 (see Theoretical Curve I). The discrepancy suggested that there was in fact an appreciable bias towards the response "old."

This was particularly evident when the over-all probability of responding "old" (regardless of whether the stimulus was in fact new or old) was plotted as a function of the mean delay for a deck. Theoretical Curve I in Fig. 3 represents the prediction from the model on the

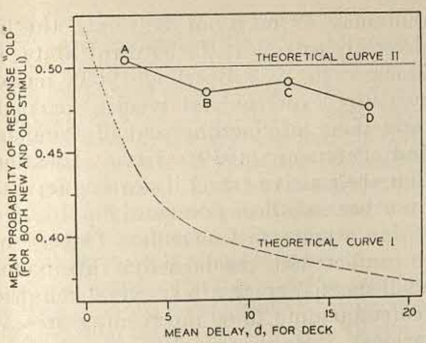


FIG. 3. Average probability of classification of both old and new three-digit numbers in a deck as "old" as a function of the mean delay between first and second presentations of numbers in that deck. (The point labeled "A" represents the mean probability for Decks A_1 and A_2 of the fixed delay condition, etc.)

assumption of no bias. But the empirical points, far from following that curve, remained more nearly in the proximity of the .5-probability line (Theoretical Curve II). There is a suggestion, therefore, that *S*s tended to readjust their response bias for each deck so as to match the objective probability of an old stimulus (which, as they could infer from the instructions, was always .5).

In accordance with this argument, when B was allowed to assume a value smaller than unity (representing a bias towards the response "old"), the theoretical curves were brought into a satisfactory fit with the data. The resulting estimates for the five parameters of the model were then as follows: $B = .42$, $w = .530$, $V = .996$, $V_c = .61$, and $u = .990$. If, as an approximation, the stimuli are regarded as falling in two classes; namely, one containing relatively distinctive three-digit numbers (such as the number 444 or *S*'s street address) and one containing relatively similar numbers (such as those in the cluster 355, 553, 335, and 533), then the parameters can be interpreted as follows (see Shepard, 1961): V represents the rate of diffusion of traces between the more distinctive numbers; V_c represents the rate of diffusion within clusters of similar numbers; w reflects the fraction of the stimuli that are more distinctive; and u specifies the rate at which traces dissipate through simple attrition.[3] When these values were substituted into the equations of the model, the curves labeled "Theoretical Curve II" in Fig. 1 and 2 were obtained. By adjusting 2½ parameters per plot, then, the achieved fit was apparently about as good as the variability of the data would allow.

Implications for informational estimates.—An expression has already been derived for a lower bound on the amount of information retained in recognition experiments in which each stimulus is presented only once and in which steady state is reached so that the probabilities $P_d(T)$ no longer depend upon T (Shepard, 1961). In the present experiment each stimulus was presented twice and the probabilities $P_d(T)$ were still increasing at the end of 200 presentations. However the trace model provides for an extrapolation to the conditions required for the informational calculations. According to the empirically determined values of the parameters of the model, the steady-state curve of forgetting is described by the equation:

$$P_d(\infty) = \{.500(.986)^d + .443(.60)^d + .057\}^B$$

And, since the traces are assumed to dissipate and diffuse independently of each other, this same equation should also apply when each stimulus is presented twice (as in the present experiment). If the informational calculations are carried out on the basis of this equation with $B = .42$, the lower bound on the amount of information carried along by S during steady state is estimated as 32 bits. However, according to the model, *S*s could as well have adopted any other bias without influencing the underlying trace process wherein the information is actually retained. Therefore, since the informational estimate is affected by B, the possibility remains that *S*s in fact carried along still more information that, owing to their response bias, was not completely reflected in their recorded responses. When B was varied in the steady-state equation, the informational estimate was found to climb rapidly from 0 (for $B = 0$), to level off at a maximum of about 50 bits near the null bias point ($B = 1$), and then to decline slowly with increasing B. This suggests that *S*s are really capable of carrying along at least 50 bits in steady state. However, in order to secure this conclusion, it would be necessary either (*a*) to drive the parameter B away from the .42-region through the experimental introduction of various payoff contingencies or else (*b*) to replace the present "old"-"new" dichotomy with a scale accommodating several degrees of confidence that the stimulus is "old" (e.g., as was done by Egan, 1958).

---

[3] By necessity, u must be less than unity; for, if there were no attrition of traces, S would eventually be swamped by the generalization from old stimuli and, thereafter, would be totally incapable of discriminating new stimuli from old. A parallel result emerged in the earlier application of the trace model to generalization in paired-associate learning (Shepard, 1958, p. 251).

In the familiar digit-span test *S*s have generally been found capable of retaining about 23 bits from a single exposure to a sequence of decimal digits (Miller, 1956; Pollack, 1953). The present informational estimates, though based on the retention of decimal digits also, are substantially higher. Possibly the use of the more sensitive recognition method (Luh, 1922) and/or the segmentation of the sequence of digits into groups of three (Martin & Fernberger, 1929; Pollack et al., 1959) is responsible for this difference. Then too, the present procedure and the digit-span procedure may tap essentially different modes of operation of the memory system. Certainly the information is retained under the present procedure in a much more diffuse form. Whereas in the digit-span test each of a small number of items must be retained with high probability, here each of a very large number of items can be retained with relatively low probability.

The memory process has previously been described in terms of fading traces that may or may not be reinforced by covert rehearsal (e.g., by Brown, 1958; Mackworth, 1959; Sperling, 1960). However the experimental results, particularly those of Sperling and of Peterson and Peterson (1959), have indicated that the passive trace, if not maintained by rehearsal, decays essentially to zero within a matter of seconds. This seems to conflict with the present evidence for a substantial trace after several minutes (corresponding to 60 intervening presentations). However, although the rehearsal of one item following the presentation of a subsequent item probably seldom occurred, each three-digit number may well have been rehearsed once or twice before passing on to the next number. Such an initial boosting of the trace might account for its subsequent persistance over greater delays.

SUMMARY

The retention of information was investigated under conditions approaching a steady state. Specifically, for each three-digit number in a protracted sequence of such numbers, *S*s responded "old" or "new" depending upon whether they did or did not remember having seen that number at some earlier point in the sequence. Each number was presented twice; but the number of other three-digit numbers intervening between the first and second presentation of the same number was manipulated as the independent variable. The results were as follows:

1. The probability of correct classification of an old three-digit number as "old" was found to be a decreasing function of the number of intervening presentations. This "curve of forgetting" dropped (from near 1.00) to about .56 for delays of 60 intervening presentations. Even after such long delays, however, the probability of correct recognition was still well above the "false alarm" probability of classifying a *new* stimulus as "old."

2. This probability of false alarm was not constant over the series of numbers but, rather, climbed rapidly (from 0.0) at first and then seemed to be slowly approaching an asymptote near .30 after about 200 presentations.

3. A trace model was shown to yield theoretical curves that could be brought into a close fit with the above two curves. The fact that the curve of forgetting departed systematically from a simple exponential

decay function was interpretable within the model as indicating that the stimuli were not all equally similar to each other. The shape of the false-alarm curve could be interpreted as indicating that *S*s tended to bias their responses towards the response "old" in such a way as to match the objective probability of an old stimulus. Both curves support the notion that traces not only fade away but, at the same time, gradually spread out so as to become decreasingly stimulus specific.

4. The fitted curves also made possible the estimation of a lower bound on the amount of information that *S*s are able to carry along at any one time during an arbitrarily protracted sequence of three-digit numbers.

REFERENCES

ANDERSON, N. S. Poststimulus cuing in immediate memory. *J. exp. Psychol.*, 1960, **60**, 216–221.

BJÖRKMAN, M. *Measurement of learning.* Stockholm: Almqvist & Wiksell, 1958.

BOURNE, L. E., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BOWER, G. H. A theory of serial discrimination learning. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959. Pp. 76–93.

BROWN, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10**, 12–21.

CAHILL, H., & HOVLAND, C. I. The role of memory in concept formation. *J. exp. Psychol.*, 1960, **59**, 137–144.

EGAN, J. P. Recognition memory and the operating characteristic. *USAF Cambridge Res. Cent. tech. Note*, 1958, No. AFCRC-TN-58-51. (AD-152650)

ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

HANAWALT, N. G. Memory trace for figures in recall and recognition. *Arch. Psychol., NY*, 1937, No. 216.

HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.

JENKINS, J. G., & DALLENBACH, K. M. Oblivescence during sleep and waking. *Amer. J. Psychol.*, 1924, **35**, 605–612.

LLOYD, K. E., REID, L. S., & FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

LONDON, I. D. An ideal equation derived for a class of forgetting curves. *Psychol. Rev.*, 1950, **57**, 295–302.

LUH, C. W. The conditions of retention. *Psychol. Mongr.*, 1922, **31**(3, Whole No. 142).

MACKWORTH, J. F. Paced memorizing in a continuous task. *J. exp. Psychol.*, 1959, **58**, 206–211.

MARTIN, P. R., & FERNBERGER, S. W. Improvement in memory span. *Amer. J Psychol.*, 1929, **41**, 91–94.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

POLLACK, I. The assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421–435.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

SHEPARD, R. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychol. Rev.*, 1958, **65**, 242–256.

SHEPARD, R. N. Application of a trace model to the retention of information in a recognition task. *Psychometrika*, 1961, **26**, 185–203.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

STRONG, E. K. The effect of time-interval upon recognition memory. *Psychol. Rev.*, 1913, **20**, 339–372.

VON FOERSTER, H. Quantum mechanical theory of memory. In H. von Foerster (Ed.), *Cybernetics: Transactions of the sixth conference.* New York: Josiah Macy, Jr. Foundation, 1950.

WILLIAMS, O. A study of the phenomenon of reminiscence. *J. exp. Psychol.*, 1926, **9**, 368–387.

WITTE, R. S. A stimulus-trace hypothesis for statistical learning theory. *J. exp. Psychol.*, 1959, **57**, 273–283.

YNTEMA, D. B., & MUESER, G. E. Remembering the present state of a number of variables. *J. exp. Psychol.*, 1960, **60**, 18–22.

(Received July 20, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 310–311

SUPPLEMENTARY REPORT: INFORMATION DEPRIVATION AND IRRELEVANT DRIVE AS DETERMINERS OF AN INSTRUMENTAL RESPONSE[1]

AUSTIN JONES

University of Pittsburgh

Some behavior, particularly of humans, appears interpretable as instrumental acts serving to reduce environmental uncertainty. In two experiments by Jones, Wilkinson, and Braden (1961), college *S*s were deprived of visual information by requiring them to lie quietly in a darkened room. In Exp. I, *S*s were allowed continuous use of a button which caused a series of lights to appear. Response frequency increased sharply over the first 9 hr., falling off slightly in the last 3 hr. In Exp. II, response frequency was an increasing function of the amount of relative information contained in the patterning of the light series. The experiments were interpreted as supporting a drive formulation of information deprivation.

The present experiment continues the assessment of an information drive concept by studying the effects of aversive and appetitive irrelevant drives upon the instrumental response of information-deprived *S*s. Hunger, and the pain associated with electric shock were selected as the irrelevant drives. The hypotheses were derived from a view of drive summation consistent with that advanced by Hull (1943). It was hypothesized that increasing proportions of control *S*s' responses would be made over time, with a slight decrement of response in the last quarter, as was found in the previous study, but that proportions of responses in the shock group would be facilitated in the initial hours of the session, then followed by a decrement over the remaining hours as a result of accommodation to the repeated presentations of the shock. (Such accommodation has been clearly shown in a conditioned avoidance procedure reported by MacDonald [1946].) It was hypothesized that *S*s entering the session under 14 hr. food deprivation would have higher proportions of responses associated with the initial hours than would control *S*s, and that proportions of responses would be an increasing monotonic function of hours, rather than decreasing in the last quarter, as predicted for control *S*s. The hypothesis of an increasing monotonic function is necessitated by the assumption that the drive stimuli associated with food deprivation increase with hours of deprivation, at least over the moderate deprivation range employed here.

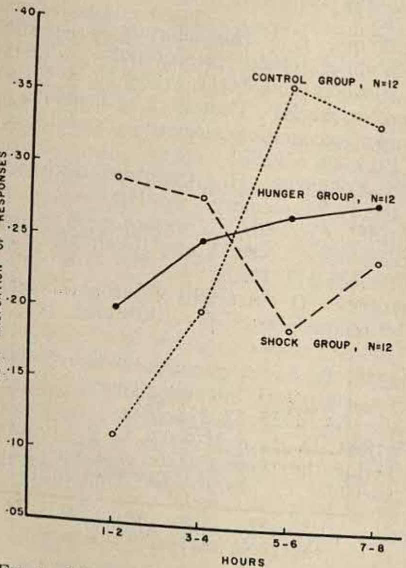


FIG. 1. Mean proportions of total responses as a function of hours in the deprivation situation.

Method.—The procedure was similar to that of the previous study. Thirty-six college *S*s served in the information-deprivation condition for 8 hr. They were divided into three groups, one receiving occasional electric shocks throughout the session, another being deprived of food beginning 14 hr. prior to the session and continuing to its end, and a control group. The *S*s were allowed to press a button causing patterned series of lights to appear, the relative information value of the series being determined by *S* through the setting of a dial.

Results.—The trends of response over time were compared by analysis of variance. The shock group and hunger group trends were each found to differ significantly from the control group trend ($P < .01$ and $P < .05$, respectively, two-tailed tests). As may be seen in Fig. 1, the differences are of the kind

---

[1] This study was supported in part by Grant M-2479 from the National Institute of Mental Health, United States Public Health Service. The author is indebted to Phillip Davis for assistance in the execution of the experiments and to Harry Fowler for comments on the manuscript.

hypothesized from the drive-summation formulation. The results are interpreted as further indication of the usefulness of an information drive concept.

A separate analysis of variance analyzed the trends of response over information values of the 36 *S*s pooled. Response was shown to be an increasing linear function of the amount of information associated with the light series ($P < .025$), thus corroborating a principal finding of the previous study.

REFERENCES

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

JONES, A., WILKINSON, H. J., & BRADEN, I. Information deprivation as a motivational variable. *J. exp. Psychol.*, 1961, **62**, 126–137.

MACDONALD, A. The effect of adaptation to the unconditioned stimulus upon the formation of conditioned avoidance responses. *J. exp. Psychol.*, 1946, **36**, 1–12.

(Received July 26, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 311–312

SUPPLEMENTARY REPORT: THE INFLUENCE OF ONE STIMULUS ON THE PREDICTION OF THE ALTERNATIVE STIMULUS IN TWO-CHOICE PROBLEMS

IRVIN RUBINSTEIN

Walter Reed Army Institute of Research

Goodnow, Rubinstein, and Lubin (1960) have shown that *S* tends to follow the conditional probabilities of a stimulus in making his predictions. The concern of this report is the effect on *S* of the conditional probabilities of one stimulus when he is predicting after occurrences of the alternative stimulus.

Method.—Two groups of *S*s, X and Y, were asked to predict a series of blue and orange cards. The length of the series for Group X was 124 trials, for Group Y it was 126. Group X was given a conventional 50:50 series randomized on both blue and orange. It was arranged so that all conditional probabilities were equal to .50 and there were no runs of either color greater than 5. For Group Y the series was 49% blue to 51% orange overall. The slight discrepancy from .50 was undesirable but unavoidable. The runs of blue for this group were exactly the same as those for Group X, not only in overall composition but also in the order in which the runs were presented. The orange cards were, however, always presented in run lengths of two for Group Y. There were 32 of these runs of orange. The conditional probabilities differed markedly from those of the orange runs given Group X.

Procedure was similar to that used by Goodnow et al. (1960). The *S*s were United States Army enlisted men. Each group made up a special type unit in the United States Army. Consequently the two groups were well matched for job and rank. There were 38 *S*s in Group X and 37 in Group Y.

Results.—The series has been split into thirds for the analysis. Since there were a total of 32 runs of each stimulus, the first third includes Runs 1–10, the second, Runs 11–21, and the third, Runs 22–32. Two methods of analysis appear feasible. First, we may compare the differences between the two groups on their prediction of blue after the occurrence of 1b (a single blue) and 2b (a double blue). This should be decisive since the only differences between their stimulus series were in the structure of the orange. Since $p(o/bo)$ for Group X is .50, while for Group Y it is 1, we expect that Group Y would be higher than Group X in its predic-

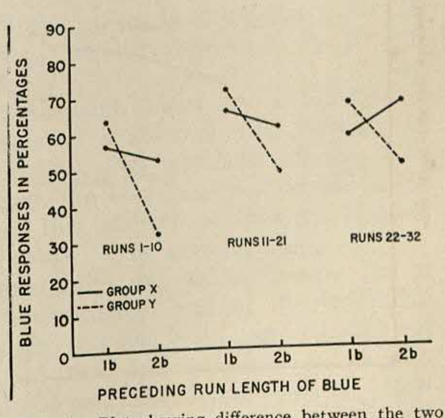


FIG. 1. Plot showing difference between the two groups in their choice of blue following one blue (1b) and two blues (2b).

tion of blue following 1b. On the other hand since for Group X p(o/boo) is .50, while for Group Y it is 0, Group Y should be lower than Group X in its prediction of blue after a run of 2b. The second approach is simply in taking the correlation in Group Y between *S*'s prediction of blue given 1b, p(b/ob), and his prediction of orange given 1o, p(O/bo), and also between p(B/obb) and p(O/boo)[1]. These conditional probabilities are unrelated insofar as the structure of the stimulus series is concerned. Any correlation would, therefore, have to be due to the tendency of *S* to react similarly to both stimuli irrespective of their differential structures.

Figure 1 shows the percentage of choices of blue for the two groups when the series is split into thirds. Since Group Y is expected to exceed Group X's prediction of blue after 1b, whereas the reverse should be true for the prediction of blue after 2b, the difference *D* between the two scores has been taken. This *D* score should maximize the effect of the experiment.

The *t* ratios between the *D* scores of the two groups for the three thirds are 3.86, 2.95, 3.67, with 73 *df*. The significance levels are .001, .01, and .001, respectively.

In Group Y, $r_{(\text{B/ob})(\text{O/bo})}$ for the three thirds are .34 ($P = .05$), .35(.05), and .31(.10). The $r_{(\text{B/obb})(\text{O/boo})}$ are .37(.05), .10(NS), and .60(.001).

The hypothesis appears to be confirmed from both approaches. The *t* ratios given are all in the expected direction. The significance levels are given from the two-tailed test, although a one-tailed test would be justified.

Discussion.—The fact that *S*s in predicting two-choice series respond to the length of run of like symbols has been demonstrated by Goodnow et al. (1960). The results of the present experiment point out, however, that *S* is also influenced by the length of run of the alternative stimulus. The *S* appears to confuse and intermix his ideas about the two separate events. This appears to be so much the case, that aside from the statistical results given in this paper, only one of 37 *S*s was able to verbalize the situation given Group Y, where the run structures of the two colors were utterly different.

REFERENCE

GOODNOW, J. J., RUBINSTEIN, I., & LUBIN, A. Response to changing patterns of events. *Amer. J. Psychol.*, 1960, 73, 56–67.

(Received January 17, 1961)

---

[1] Upper case letters refer to responses, lower case to stimuli.

Journal of
Experimental Psychology

Vol. 62, No. 4 October 1961

SOME FACTORS IN THE PROGRAMING OF CONCEPTUAL LEARNING[1]

ROBERT M. GAGNÉ AND LARRY T. BROWN

Princeton University

Studies of conceptually mediated behavior in human beings are often concerned with the effects of variations in previously established concepts on behavior in the solution of problems. This is true, for example, of the work of Maier (1930) with practical problems; of Luchins (1942) on water-jar problems; and of Maltzman and his collaborators (Malzman, Eisman, Brooks, & Smith, 1956; Maltzman & Morrisett, 1952, 1953a, 1953b) on anagrams; to mention a few well-known investigations. Usually, in such studies, the concepts employed by *S* are assumed to have a history of previous establishment, and are considered to be available to *S* at the time the problem is set. In a different category can be placed research on the learning (or formation) of concepts, such as that of Hull (1920), Smoke (1932), Reed (1950), and Heidbreder (1947), which need not be reviewed here. The concepts acquired in the course of the experiment are usually not further "used," as in the solution of a problem, but are simply measured as being "established" in the sense that they meet a criterion of learning or recall.

It would appear worthwhile to devote some effort to the experimental study of events which bridge these two processes of concept learning and utilization; in other words, to examine the question of how concepts (or concept sequences) which are newly learned subsequently enter into the activity of solving problems. There is the related question, too, of the extent to which observations of problem solving performance may throw light upon the effectiveness of learning of the concepts employed. This is the general framework of the present study.

Although there are a number of isolated studies of the learning of principles for problem solving, perhaps the best known intensive work on this subject is that of Katona (1940). His investigations demonstrate the relative ineffectiveness of "memorizing" and of "verbal principle learning" in the solution of card-trick and match-stick pattern problems, as compared with what Katona refers to as "understanding."

---

[1] The research reported in this paper was supported in part by a grant from the Carnegie Corporation of New York.

A number of Katona's findings have been confirmed and amplified by other investigators (Hilgard, Edgren, & Irvine, 1954; Hilgard, Irvine, & Whipple, 1953). According to Katona, the learning of "senseless connections" may be contrasted in its effectiveness for the solution of new problems with the learning of "meaningful organization." When learning partakes of the apprehension of meaningful wholes, he considers it to be the sort which favors transfer to problem solving situations. In Katona's experiments, the greatest success in solving card-trick problems was achieved by *S*s who listened to an explanation of a basic problem and watched a step-by-step demonstration on the part of *E*. In the case of these materials, this method was found to be the most effective, followed in order by (*a*) learning a verbally stated principle; (*b*) memorizing the steps in the procedure; and (*c*) no training.

Of considerable interest to those who want to study conceptual learning has been the recent development of teaching machines and the programing of learning materials for such machines (see Galanter, 1959). In general, descriptions of these developments have made it clear that the intended use, at least, of teaching machines and their related procedures is for the establishment of useful concepts (often called "verbal behavior"), as opposed to the memorization of rote materials. It would also appear that most investigators of learning programing believe that procedures can be developed which will convey the kind of "understanding" that Katona describes, at least in terms of their effectiveness in establishing whatever it takes to solve problems. However, from an examination of representative published examples of programs (e.g., Holland, 1959; Skinner, 1958) it is not immediately apparent that they are conveying "understanding" in the sense of capability for inducing transfer to new problem situations. They appear to be concerned primarily with the usage of words in a variety of stimulus contexts (e.g., the word "incandescent" in Skinner's 1958 example). It seems possible, therefore, for such programing of conceptual material to induce the learning of "verbally-stated principles" which Katona considers from his evidence to be less than optimal for problem solution. Whether this is actually the case or not cannot be told from present evidence. The very important reason for this is that no investigation of learning programing has used a measure providing evidence of transfer of the learning to performance in a problem situation.

Specifically, the purpose of the present study was to obtain a measure of learning effectiveness of programed material, in terms which would permit a reasonable inference that "understanding" had been accomplished, and to relate this measure to certain characteristics of the programing. The materials pertained to number series, and it was intended that *S*s would learn to state and use formulas for the sum of any number of terms in such series. In measuring effectiveness, a test was used which required transfer of learned principles to a novel problem solving situation. The characteristics of the program which were varied may be described as follows: (*a*) step-wise vs. abrupt presentation (an extreme of "large steps" in a program); (*b*) encouragement of "discovery" of principles vs. "verbal statement" of principles. It may be said parenthetically that the relatively inadequate specificity with which such variations in stimulus materials can presently be described is fully recognized. It was hoped that the present experiment would point the way to further refinement of such description in research to follow.

METHOD

Apparatus

For the presentation of learning materials, a simple form of a teaching machine was employed. Cards (4 × 6 in.) containing the items to be responded to were placed in individual plastic cases of a visible card file. Each file contained 25 items. The fiberboard backing of the card file was then clipped to a board mounted on a stand inclined at about 40°, which was placed on a table in front of *S*. The correct response to each item was printed on the back of the card. The *S* read the item, wrote his response on an answer sheet numbered to correspond with the item numbers, and then flipped the card downward towards him to check the answer shown on the back of the card. When 25 items were completed in this way, *E* simply replaced the visible card file with another containing subsequent items in sequence.

Materials

The learning programs which were devised to teach principles pertaining to number series consisted of an Introductory program, used by all *S*s; and three programs incorporating differences in experimental treatment, which were designated Rule and Example (R&E), Discovery (D), and Guided Discovery (GD). The Introductory program contained 89 items, beginning with one which read: "Here is a series of numbers: 1 3 5 7 9 11 13 — —. What are the next two numbers in this series? — —." Its intention was to establish learning of basic concepts, in a definitional sense, that would be used in later learning and problem solving. Specifically, these concepts were *term value* (any individual term in the series), and its symbol, T; *term number* (the position of a term in a series) and its symbol, n; and other derivatives of these such as $n+1$, T_{n-1}, Σ^{n+1}, etc. The program was constructed of small steps, in accordance with principles described by Skinner (1958, 1959) and various other investigators (cf. Galanter, 1959). It was revised after initial tryout with a few *S*s, and seems to have achieved a fairly successful form. (The average number of errors made by boys of Grades 9 and 10 over the 89 items was 12.3.)

Each of the experimental treatment programs began with the item: "Here is another series. Series: 1 2 4 8 16 — —. Now we want to obtain a formula for Σ^n for this series. First, can you fill in the next two term values? — —." Each then provided, in the different ways to be described, practice in using the basic concepts previously learned in connection with the following four number series:

(1) 1 2 4 8 16 — —
(2) 1 3 5 7 9 11 — —
(3) 1 2 4 8 16 — —
(using a different formula)
(4) 1 3 7 15 31 — —

In each case, the series was ultimately displayed in three rows which identified the term number, the series, and the sum of terms. In addition, arrows were used to direct attention to the column in which the terms to be employed in the formula were to be found. The following is an example:

$$\begin{array}{rccccccc}
\text{Term No.:} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\text{Series:} & 1 & 2 & 4 & 8 & 16 & 32 & — \\
 & & \nearrow & \nearrow & \nearrow & \nearrow & \nearrow & \\
\Sigma: & 1 & 3 & 7 & 15 & 31 & — & —
\end{array}$$

The R&E program, following one or two introductory items for each number series, began by stating the correct formula in each case, which *S* was instructed to write down on his answer sheet. (For example, for Series 1 the formula stated was $\Sigma^n = T_{n+1} - 1$.) Following this, items progressed in small steps through a number of examples which required the identification of terms in the formula, and the finding of numerical values for Σ^n by the use of the formula. This program contained Items 90 − 125, or a total of 36.

The D program began with the same introductory items, then abruptly stated "What is the rule for Σ^n, in terms of any term number (n) or term value (T), or both? Try to do this by yourself. If you need a hint, turn to the next card. If you know the answer, TELL ME." The following cards contained hints which further directed the attention of *S* to certain relationships between a specific sum such as 7, and a term value such as 8, by outlining these in red ink. Questions contained on each successive "hint" card gave progressively more complete guidance, without, however, stating the answer. It was necessary for *S* to show his answer to *E*, since there are several correct formulas for each series. As a whole, then, the D program had the following sequence: Number Series 1: 3 items, 2 hints; Number Series 2: 2 items, 3 hints; Number Series 3: 1 item, 6 hints; Number Series 4: 1 item, 3 hints.

The GD program began with the same introductory items as the others. Following this, an item stated: "What do you have to do to obtain the sum, or Σ? Can you get Σ^n by adding, subtracting, dividing, or multiplying the numbers in the row pointed to? For example, when Σ^n is 3 (bottom row), the term-value pointed to is 4. If you know 4, you can get 3 by ——— ———?" Successive items gave further examples of this relationship, in a small-step fashion, until an item was reached which read: "Are you ready to state a general rule? Try it in symbols, letting T_{n+1} mean the term-value of the $n+1$st term." Then followed several examples which required the use of the formula in finding specific numerical values for Σ^n. Altogether this program contained 40 items, numbered from 90 to 129.

For all three groups, the same problem solving test was used to measure performance. This test comprised four problems, administered individually, each of which involved a number series with which *S*s had no previous acquaintance. Each of the series was shown in a center row labeled "Series," along with an upper row of digits labeled "Term No." and a lower row labeled "Σ." In each case, the instructions called for *S* to find the formula for Σ^n. The series employed in the test were these:

Problem 1:	1	3	9	27	81	—	—
Problem 2:	0	2	4	6	8	10	—
Problem 3:	½	1	1½	2	2½	3	—
Problem 4:	1	2	3	4	5	6	—

The problems were also accompanied by a set of hints to be displayed successively in a visible card file, under conditions to be described. These were constructed similarly to those hints used in the D program.

Subjects

The *S*s were boys in Grades 9 and 10, ranging in chronological age from 14 to 16 yr. who had no previous specific knowledge about number series. Nearly all these boys had either completed or were completing a first course in algebra; a few were studying general mathematics which does not include algebra. Their grades in Grade 9 mathematics covered a range from D+ to A. Equivalent IQ scores on the Henmon-Nelson intelligence test ranged from 100 to 154, indicating a very considerable spread of basic academic ability. Eleven *S*s were assigned at random to each of the three experimental groups, for a total of 33. The administration of the program was discontinued in the case of 2 additional *S*s who made more than 55 errors on the Introductory program, and were therefore assumed to be quantitatively not comparable to the other *S*s.

Procedure

In broad outline, the procedure began with the giving of instructions and the administration of the Introductory program of 89 items. After a 3-min. break, *S* was assigned at random to one of the three experimental groups, and proceeded to complete one of the three learning programs. He returned the following day at the same hour, and went through the same program (excluding the Introductory) once again. A 3-min. break was given, and *S* received instructions for the problem solving test, which he then took.

Introductory program.—Instructions were given to the effect that we wanted *S* to learn some things about number series. He was then told how to respond to the cards and how to manipulate them. In the case of an incorrect answer, he was to turn back the card, read it again, cross out his answer, and write the correct one, before going on. The *E* kept a record of errors and recorded time to complete the 89 items. Very infrequently, he answered questions by repeating relevant parts of the instructions.

Learning program.—The items of one of the three learning programs were administered in the usual way, following the 3-min. rest. Again, *E* recorded errors and time for completion. With the D program, he also recorded the number of hints used by *S*. The procedure employed with the three learning programs insured that, although the *S*s responded to different sets of learning materials, they all achieved successful answers to questions about the *same four number series.* To make doubly sure that this was so, each *S* went through the same learning program a second time on the day following his first session. In this respect, it may be said that all *S*s had roughly the same degree of previous experience with number series at the time of the test.

Problem solving test.—After a 3-min. rest, instructions were given for the test. These stated that the problems would be given one at a time, and 10 min. allowed for completion of each problem. The *S* was told to keep track of time by glancing at an electric timer placed nearby. He was instructed to try to solve the problem (finding a formula for Σ^n)

by himself. At the end of 5 min., however, *E* would say "Look at Hint Number 1," at which time *S* was required to turn a card exposing a hint. He could use further hints if he felt he needed to, without additional instructions from *E*.

The test was designed to obtain maximal measurement potential from a limited number of number series available (since there are not many which can be treated in the manner covered by the learning program). The *E* kept a record of: (*a*) time to solve each problem (a problem not completed was recorded as requiring the maximum time of 10 min.); (*b*) number of hints employed; and (*c*) incorrect answers. The last-named measure yielded few usable scores, and was not employed in the analysis of results. Since each series can yield more than one correct formula, *S* was instructed to show his answers to *E*, who would tell him whether they were correct. The time at which *S* showed an answer was noted and recorded if the answer was correct; otherwise the timer was kept running and *S* told to keep trying. In case the problem was not solved at the end of 10 min., efforts were made to prevent discouragement of *S* by pointing to the numerals involved in the correct formula, without, however, expressing these in words or symbols. As for correct answers, verbal expressions (using the phrases "term number," "term value," etc.) were accepted in place of symbolic expressions (using n, T_n etc.), although these were not frequent.

Scoring performance.—The available measures of performance on the problem solving test were time taken to solve, and number of hints employed. In addition to these, a proficiency measure was constructed which utilized both kinds of information, namely, a weighted time score, using the formula:

$$\text{Weighted Time Score} = \text{Min.} + 4(\text{Hint } 1) + 3(\text{Hint } 2) + 2(\text{Hint } 3) + 1(\text{Hint } 4)$$

Subject to the possible distortion arising from combining experimental groups, the split-half reliability of this score was determined to be 0.72.

RESULTS

Introductory program.—Time to complete the introductory program ranged from 26.8 to 53.9 min., and errors on these 89 items from 2 to 37. Using a score of time and errors, a comparison was made among those assigned to the three experimental groups. The results were as follows: Cond. R&E, *M* = 51, *SD* = 13; Cond. D, *M* = 49, *SD* = 12; Cond. GD, *M* = 53, *SD* = 10. So far as this initial performance measure is concerned, the groups may be assumed to be comparable.

TABLE 1

MEANS, *SD*s, AND MEAN DIFFERENCES OF TIME, ERRORS, AND HINTS DURING FIRST AND SECOND LEARNING SESSIONS, FOR EACH CONDITION

Measure	Condition					
	R&E ($N = 11$)		GD ($N = 11$)		D ($N = 11$)	
	Time	Errors	Time	Errors	Time	Hints
1st Learning						
Mean	28.8	6.2	33.2	10.1	18.5	3.9
SD	8.7	4.4	6.9	4.4	5.5	1.6
2nd Learning						
Mean	15.4	3.5	23.3	7.1	9.8	2.5
SD	3.4	3.2	6.6	3.7	5.3	2.0
Difference in Means	13.4	2.7	9.9	3.0	8.7	1.4

Learning program.—Table 1 provides values of means and *SD*s for the time taken to complete learning during the first and second (relearning) sessions. Errors during these sessions are given for Cond. R&E and GD, and number of hints required for Cond. D. Comparisons among the groups in learning scores simply reveal differences in length (or difficulty) of the three learning programs. It is of some interest to note that Cond. D requires relatively the shortest time, during both the first and second learning sessions. The two other learning methods are more nearly comparable to each other in this respect, with Cond. GD showing greater time and error scores.

The main comparison to be made is between the scores for the first and the second learning sessions, within each condition. The data of Table 1 indicate reductions in both time and errors when these sessions are compared. The differences for errors and for hints (Cond. D) are not significant. Differences in time scores between the two sessions are significant beyond the .01 level (t test) for all three groups. Using the time scores as measures of learning leads to the conclusion, therefore, that significant learning gains took place with all three of the experimental learning methods.

Problem solving performance.—The results of administering the test containing four new number-series problems are contained in Table 2, showing means and *SD*s for the three measures used. Analyses of variance performed on these data yielded the following values: for time scores, $F = 6.77$, $P < .01$; for number of hints, $F = 12.58$, $P < .01$; for weighted time scores, $F = 13.44$, $P < .01$. In the case of time scores, Bartlett's test indicates homogeneous variance; for the other two measures, however, the variances are significantly heterogeneous. Tests of t between individual means yielded the following values for the time scores: Cond. R&E vs. Cond. D, $t = 5.29$, $P < .01$; Cond. R&E vs. Cond. GD, $t = 9.26$, $P < .01$; Cond. D vs. Cond. GD, $t = 2.48$, $P < .02$. For number of hints, the values for these comparisons in the same order were $t = 8.10$, $P < .01$; $t = 9.31$, $P < .01$; $t = 1.21$, $P > .20$. For the weighted time scores, these values were $t = 7.61$, $P < .01$; $t = 9.96$, $P < .01$; and $t = 2.36$, $P < .05$.

TABLE 2

MEANS AND *SD*s OF THREE PERFORMANCE MEASURES ON THE FINAL NUMBER-SERIES TASKS, FOR EACH CONDITION

Cond.	Measure					
	Time (Min.)		No. of Hints		Weighted Time Score	
	M	*SD*	*M*	*SD*	*M*	*SD*
R&E	27.4	7.1	6.5	4.0	46.8	16.8
GD	16.7	3.5	1.5	1.1	22.6	7.2
D	19.8	6.1	2.2	1.0	28.4	7.5

In sum, both the time scores and the weighted time scores show significant differences (.05 level) between means for the individual conditions. In the case of number of hints, the means of Cond. D and GD do not differ significantly, but both differ from the mean of Cond. R&E. The most effective condition of the three appears to be GD; R&E is the least effective, and D somewhere in between.

DISCUSSION

Several things are shown by the results of this experiment. First of all, it is clear that the construction of learning programs by the use of small steps and the manipulation of other, similar variables which obliquely characterize content, can lead to programs which differ markedly in effectiveness. By the latter term we mean here the transfer of what is learned to a problem solving situation. In this study, both the R&E program and the GD program were designed to provide the kind of learning situation which Skinner (1958) and others have described. (It is recognized, however, that a further reduction in errors would be possible by redesign of some items.) The results obtained by comparison of the first and second (relearning) sessions give evidence that a significant amount of learning occurred in both programs. Just as obviously, learning also occurred in the D program, which may in one sense be characterized as having used "very large steps." Although of course no direct comparison is possible because of the differences in these materials, it would apparently be difficult to make a choice among these programs, insofar as amount of learning is concerned. All of them show significant gains in performance from the first to the second learning session. They are all comparable in the sense that *S*s "worked through" the same four number series.

When a final performance in finding formulas for four new number series is measured, however, the picture is a very different one. With such a transfer measure, the performances of the three groups split apart quite markedly. The programs containing small steps are the ones which differ most in their effectiveness in terms of this measure, whereas the large step program (D) falls in between. The suggestion is irresistible that some aspect of *what* has been learned in this situation is of greater effect than *how* it has been learned.

The finding that a Discovery method leads to greater transfer than does a Rule and Example method is quite consistent with the findings of Katona (1940) as well as other investigators (Haslerud & Meyers, 1958; Hilgard et al., 1954; Hilgard et al., 1953). Katona's work itself suggests that the individuals who "memorize" solutions to his problems were simply not learning the right things. Left to discover solutions, they somehow were able to acquire the "knowledge" needed to solve new and unfamiliar problems. However, a noteworthy aspect of the present findings is that it was possible to guess at what some of these right things might be, and so to construct a Guided Discovery program which was more effective than either of the other two.

One of the limitations of this study, certainly, is that it does not tell us exactly what these "right things" are. It does not seem possible, at the present state of our knowledge, to design an experiment in this field which possesses the relatively exact degree of stimulus definition characteristic, say, of experiments using nonsense syllables with known association values. Nevertheless, it may be possible to draw some inferences concerning differences between the R&E and GD programs, which will point to the kinds of variables that could be studied in more exact fashion in follow-up experiments.

What were the crucial differences between the R&E and GD programs? Both provided for step-by-step learning in connection with the same four number series. Both provided examples in which formulas for the sum of terms had to be used with specific numerical values. The difference between them appears to reside in the occurrence of the type of item described for the GD program (see Method) which does not occur in the R&E program. Examination of this item, together with one or two subsequent items of a similar nature, shows that they required the use of previously learned concepts in a new stimulus context. In responding to such items, S presumably had to actively produce such concepts as term number, n, term value, T, T_{n+1}, which he had just previously learned in the Introductory program, as well as some concepts learned even earlier, such as subtract, divide, multiply, and add. That is to say, the items specifically required S to respond by writing, or otherwise reinstating, these verbal responses. In contrast, it may be noted that the R&E program did not require this; instead, entities like term number, T, T_{n+1} occurred as stimuli (in the formula provided) to which the required responses were the locating and copying of specific numerals.

Presumably, the D program also required the use (i.e., supplying by S himself) of previously learned concepts, in a manner similar to GD. We cannot, of course, identify these specifically, as we can in the individual steps of the GD program. In other words, we cannot be sure that they were systematically practiced as responses in arriving at the formulas required as answers to the four number series. Some of them probably were, while some perhaps were not. The suggestion is, therefore, that the GD program led to a superior performance because it required systematic reinstatement of learned concepts, whereas the D program permitted such reinstatement to be more of a random or chance affair. We do not consider that our experiment has proved this, but rather believe it is the kind of inference which may reasonably be made.

It is noteworthy that the differences produced by the GD and the R&E programs resulted from differences in a relatively few items (four for each number series of the learning program). There is consequently a suggested relation between these findings and those of Maltzman and his associates (Maltzman et al., 1956; Maltzman & Morrisett, 1952, 1953a, 1953b) who have found that brief verbal instructions and brief periods of training may bring about significant differences in performance in the solution of anagram problems. In terms of Maltzman's (1955) theoretical formulation, it might be argued that a few crucial items in the GD program have served to raise the habit strength of certain connections, thus heightening the strength of essential habit family hierarchies of the sort mentioned previously (i.e., those containing such responses as T_{n+1}, n, as well as subtract, etc.). The ideas of Saugstad (1955, 1957) regarding the necessity for availability of concepts in problem solving appear also to be related. In our experiment, it seems likely that the GD program had the effect of increasing the availability of the concepts term value, T, $n+1$, etc., as well as subtract, divide, etc., as we have previously noted. However, the present results do not make it possible to decide whether the reinstatement of these concepts accomplished the strengthening of connections, or whether it perhaps aroused some previously acquired organization.

In a more general vein, it is certainly not a new idea that positive transfer of training to a problem solving situation should depend upon the resemblance between what is practiced in a learning situation and what is used in problem solving. Surely our experiment provides another demonstration of this effect; as we have stated, it appears that concepts used in finding formulas in the final performance test were most systematically practiced in the GD program. But this general rule is not so completely obvious when one tries to apply it to the development of an effective learning program. For example, it might be said that the D program, in which *S*s were required to find formulas for new number series, was most like the final situation, in which the finding of formulas for new number series was again required. Yet this was not the most effective program, in terms of positive transfer. Thus the present findings emphasize that the sources of positive transfer are to be sought in the mediators of behavior, rather than in the behavior itself. To find the ways of developing effective learning programs, one must search out the specific concepts (verbal responses, if one prefers such language) which enter into the chain of events *between* the stimulus situation and the overt performance itself. One can probably do this on an empirical basis; however, it would seem desirable to attempt a more generalizable result, by continuing the effort to obtain increasingly exact definitions of "what is learned" in the direction suggested by this study.

We consider that Katona's (1940) emphasis on "understanding" is correct insofar as it emphasizes the necessity for measuring learning effectiveness in terms of transfer to a problem solving situation. Such a method was employed in this study. However, our results emphasize the importance of "what is learned" rather than "how it is learned" as the crucial factor in learning effectiveness. Discovery as a method appears to gain its effectiveness from the fact that it requires the individual learner to reinstate (and in this sense, to practice) the concepts he will later use in solving new problems. To the extent that the GD program was able to identify these concepts, it could then provide systematic practice in their use, and thus lead to a performance superior to that attained otherwise. The practice provided by the R&E program, in contrast, did not *require* the use of these essential concepts (although it permitted it). Accordingly, it led to a distinctly inferior problem solving performance.

SUMMARY

An experiment was performed to investigate the effects of certain variations in the programing of conceptual learning materials on effectiveness of learning as measured by performance in a problem solving situation. The programs were designed to foster learning of concepts to be used in the deriving of formulas for the sum of terms in unfamiliar number series.

Thirty-three boys in Grades 9 and 10 were

assigned randomly to three groups of 11 each, and were tested individually. After first learning basic concepts pertaining to number series in a "small step" Introductory program of 89 items, the groups respectively went through three different learning programs called Rule and Example (R&E), Guided Discovery (GD), and Discovery (D). In a second session the following day, relearning of the program was followed by a task containing four problems, each requiring *S* to find the formula for the sum of *n* terms in a number series. Performance was measured in terms of time to solve, number of hints required, and a weighted time score combining these.

The results show significant learning gains ($P < .01$) between Learning Sessions 1 and 2 under each condition. Final performance scores indicated best performance for Cond. GD, worst for Cond. R&E, intermediate for Cond. D. Differences in these measures were tested by analyses of variance yielding $P < .01$ in all cases. Variances were found to be homogeneous for time scores, not for the other two measures. For all measures, the means for Cond. R&E differed significantly from means of the other conditions, when tested individually ($P < .01$). For time scores, comparisons of all individual means showed significant differences ($P < .02$).

Discussion of the results emphasizes the importance of "what is learned" as opposed to "how it is learned" for problem solving performance. The Guided Discovery program, using small steps, requires *S* to reinstate (i.e., actively produce) certain concepts, a feature which may be lacking in the Rule and Example program. The Discovery program, containing "large steps" may produce such reinstatement in a less systematic manner.

REFERENCES

GALANTER, E. (Ed.) *Automatic teaching: The state of the art.* New York: Wiley, 1959.

HASLERUD, G. M., & MEYERS, S. The transfer value of given and individually derived principles. *J. educ. Psychol.*, 1958, **49**, 293–298.

HEIDBREDER, E. The attainment of concepts: III. The process. *J. gen. Psychol.*, 1947, **24**, 93–138.

HILGARD, E. R., EDGREN, R. D., & IRVINE, R. P. Errors in transfer following learning with understanding: Further studies with Katona's card-trick experiments. *J. exp. Psychol.*, 1954, **47**, 457–464.

HILGARD, E. R., IRVINE, R. P., & WHIPPLE, J. E. Rote memorization, understanding, and transfer: An extension of Katona's card-trick experiments. *J. exp. Psychol.*, 1953, **46**, 288–292.

HOLLAND, J. G. A teaching machine program in psychology. In E. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: Wiley, 1959. Pp. 69–82.

HULL, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**(1, Whole No. 123).

KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

LUCHINS, A. S. Mechanization in problem solving: The effect of Einstellung. *Psychol. Monogr.*, 1942, **54**(6, Whole No. 248).

MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–143.

MALTZMAN, I. Thinking: From a behavioristic point of view. *Psychol. Rev.*, 1955, **62**, 275–286.

MALTZMAN, I., EISMAN, E., BROOKS, L. O., & SMITH, W. M. Task instructions for anagrams following different task instructions and training. *J. exp. Psychol.*, 1956, **51**, 418–420.

MALTZMAN, I., & MORRISETT, L., JR. Different strengths of set in the solution of anagrams. *J. exp. Psychol.*, 1952, **44**, 242–246.

MALTZMAN, I., & MORRISETT, L., JR. The effects of single and compound classes of anagrams on set solutions. *J. exp. Psychol.*, 1953, **45**, 345–350. (a)

MALTZMAN, I., & MORRISETT, L., JR. Effects of task instructions on solution of different classes of anagrams. *J. exp. Psychol.*, 1953, **45**, 351–354. (b)

REED, H. B. The learning and retention of concepts: V. The influence of form of presentation. *J. exp. Psychol.*, 1950, **40**, 504–511.

SAUGSTAD, P. Problem-soving as dependent on availability of functions. *Brit. J. Psychol.*, 1955, **46**, 191–198.

SAUGSTAD, P. An analysis of Maier's pendulum problem. *J. exp. Psychol.*, 1957, **54**, 168–179.

SKINNER, B. F. Teaching machines. *Science*, 1958, **128**, 969–977.

SKINNER, B. F. The programming of verbal knowledge. In E. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: Wiley, 1959. Pp. 63–68.

SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**(4, Whole No. 191).

(Received August 18, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 322–328

THE EFFECTS OF CONGRUENT AND CONFLICTING SOCIAL AND TASK FEEDBACK ON THE ACQUISITION OF AN IMITATIVE RESPONSE[1]

JOHN T. LANZETTA AND VERA T. KANAREFF

Center for Research on Social Behavior, University of Delaware

Tulane University

Several earlier studies (Kanareff & Lanzetta, 1958, 1960; Lanzetta & Kanareff, 1959) have demonstrated the attenuating effects of negative sanctions for imitation, induced by instructional variations, on the acquisition of an imitative response. Subjects, when faced with a choice between employing a "disapproved" imitative response which has a high probability of leading to a correct decision or utilizing a positively sanctioned response of independence or opposition which has a low probability of leading to a successful choice, appear to compromise: the proportion of imitation responses elicited is less than expected either on the assumption that *S*s attempt to maximize their frequency of success or that they "event match" (Estes, 1957).

The use of instructions to manipulate expectations of approval or disapproval, however, did not allow evaluation or control of the intensity of the applied sanctions and consequently did not provide a test of the relative effects of social and task feedback on the acquisition of an imitative response. In an effort to provide more continuous and precise control a recent study (Kanareff & Lanzetta, 1960) employed a verbal response from *E* as a social reinforcer in conjunction with an independently varied task reinforcement (indication of correctness or incorrectness). Verbal reinforcers have been shown to be effective in the conditioning of varied responses (e.g., Buss, Gerjuoy, & Zusman, 1954; Greenspoon, 1955) but they proved singularly ineffective in this study. Several possible reasons for the poor conditioning were recognized, some of which provided the rationale for the present procedure. For one, *E* had stressed the importance of being correct and *S* was provided a visual signal as an indicator of performance adequacy. Thus, *E*'s verbal comments may have evoked minimal attention. Secondly, the verbal statements "okay" and "good" may not have had sharply discriminable "meaning" as social reactions. In view of this it appeared desirable to have the social reinforcement independent of *E* as well as independent of the task reinforcement and also to utilize as socially reinforcing stimuli cues which would be more sharply discriminable. One such reinforcement would be a co-worker's emotional response to *S*'s behavior on each trial.

The present study employs the "galvanic skin response" (GSR) purportedly from a partner as a social reinforcer, in conjunction with a task reinforcement. The *S*s were given

---

[1] This research was supported by Grant NSF-G 11427 of the National Science Foundation and Grant M-3909 from the National Institute of Mental Health, United States Public Health Service. The writers wish to express their appreciation to Jo Anne Davis for her assistance in collecting and analyzing the data.

the task of predicting whether a red or green light would be illuminated. On each trial they were presented with two cues, an indication of correctness of their choice and an indication of the reaction of their partner. The *E* controlled both cues and therefore was able to control their congruence or incongruence, e.g., imitation of the partner could lead to a correct judgment but a negative emotional reaction from the partner. It was anticipated that congruence of social and task feedback would result in a higher level of imitation than when the reinforcements were in conflict. More specifically, for both task reinforcement conditions, it was predicted that the level of imitation would be a positive monotonic function of the probability of a positive social response from the "partner" for imitation.

METHOD

Apparatus and procedure.—The *S*s were tested in pairs, but were seated in separate enclosed booths. Each booth contained a response box on which were two response keys and two pairs of red and green signal lights mounted one above the other. The task was to predict whether the red or green light of the top pair of lights would be illuminated. The *S* indicated his prediction by pressing the appropriately labeled response key. Each *S* was led to believe that his partner's choice, indicated by the lights above the response keys, was given first. The *E* was thus able to simulate the partner by presenting a programed sequence of "partner's" choices to each *S*. The correctness of *S*'s choice was indicated by the top row of lights on the response box.

In addition to the response box, there was placed in each booth a "galvanometer" which *S*s were told would measure GSR reactions. Two electrodes purportedly connected to a galvanometer in the "partner's" booth were attached to *S*'s left palm in order to allow the "partner" to see *S*'s GSR. The galvanometer in *S*'s booth was similarly to inform *S* of the "partner's" emotional responses. The galvanometers were disguised dc voltmeters which were controlled by *E*. The "partners'" emotional response was programed by *E* according to the schedule of social reinforcement in effect.[2]

The sequence of task events is as follows: (*a*) A buzzer signal indicates the start of a trial. (*b*) Within 3 sec. *E* selects the "partner's" choice and it is displayed on *S*s' response boxes. (*c*) This serves as a signal for both *S*s to make a choice. (*d*) Two seconds after *S*s respond the GSR reading is displayed. It remains displayed for 4 sec., until the red or green "outcome" light is illuminated. (*e*) The "outcome" signal remains displayed for 3 sec. One second after it terminates the buzzer is sounded starting a new trial.

Instructions.—The following paraphrased portions of the instructions indicate the rationale provided to *S*s:

> Most of you are probably familiar with gambling phenomena and the kinds of guesses that people make when gambling. We're primarily interested in seeing how autonomic reactions, that is, emotional reactions are related to the types of guesses people make. You will be asked to make a choice as to whether the red or green light on this box will be illuminated. Indicate your choice by pressing the appropriate button on the box in front of you. The apparatus is constructed so as to automatically produce a sequence of red and green lights for you to predict.
>
> The particular autonomic function we'll be measuring today is palmar sweating. In the past our *S*s have shown a great deal of curiosity about this response so we tried to arrange the equipment to allow you to watch the GSR readings. Since the guesses you make might be affected if you see your own reactions, we can't let you watch your own GSR. However, we can let you watch your partner's reactions so that you can get some feeling for the way emotional tension changes over time.
>
> We have a pair of electrodes and a galvanometer. Changes in skin resistance are picked up by the electrodes and transmitted to the galvanometer. As you can see, when the pointer falls here on the scale the emotional state is positive, the person feels pleased about what goes on; when it falls in this area, the emotional state is

---

[2] We are indebted to members of the staff of the Communications Social Science Research Department of Bell Telephone Laboratories for suggesting this experimental technique.

negative, he feels displeased about what goes on. The galvanometer will be connected for readings as soon as both of you make your predictions. The pointer will rest in the center position until both of you make your predictions, it will then move to the right or left of center showing your partner's reaction to what went on.

Subjects and design.—The three probabilities of social reinforcement used were .50 for imitation, and 1.00 for opposition. On the 1.00 schedules the "partner's" emotional response was positive (pleased) for all imitation (or opposition) responses, i.e., for all choices identical (or opposite) with the partner's. On the .50 schedule the "partner's" emotional response was positive on 50% of the trials and negative on the other half, the order of positive and negative responses being randomized.

The two probabilities of task reinforcement used with each schedule of social reinforcement were .80 and .50. On the .80 schedule the "partner's" choice was called correct on 80% of the trials, i.e., imitation of the partner was instrumental to being correct. On the .50 schedule the partner's choice was correct on one-half of the trials.

The three probabilities of social reinforcement and two probabilities of task reinforcement formed a 3 × 2 factorial design. Forty-eight male and 24 female college student volunteers were randomly assigned in like-sex pairs to the six cells with the restriction that there be 8 male and 4 female *S*s in each cell.

The dependent variable was the number of imitation responses per 10-trial block. There were 100 trials.

RESULTS

Experiment 1.—Figure 1 and Table 1 indicate the major results: (*a*) The task reinforcement significantly affected the rate and level of acquisition of imitation responses. (*b*) The level of social reinforcement had little effect on the frequency of imitation for either of the task reinforcement conditions.

Interviews at the termination of the session indicate that *S*s were generally aware of the partner's emotional response throughout the session.

For the present population of *S*s and present conditions, it appears that the efficiency of a response for achieving success is more important in determining its rate of utilization

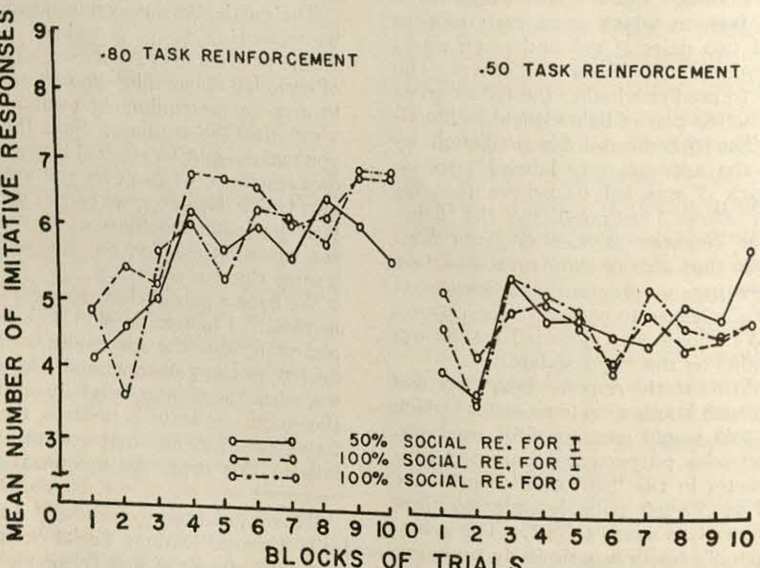


FIG. 1. Mean number of imitative responses per block of trials as a function of probability of task reinforcement and probability of social reinforcement.

TABLE 1

Analysis of Variance of Imitative Responses

Source	*df*	*MS*	*F*
Task reinforcement (TR)	1	192.20	16.15*
Social reinforcement (SR)	2	4.42	—
Trials (T)	9	16.01	9.54*
Sex (S)	1	25.07	2.11
TR × SR	2	6.43	—
TR × T	9	9.15	4.79*
TR × S	1	1.60	—
SR × T	18	2.29	1.20
SR × S	2	19.58	1.64
T × S	9	2.23	1.17
TR × SR × T	18	1.85	—
TR × SR × S	2	31.64	2.66
TR × T × S	9	1.53	—
SR × T × S	18	1.58	—
TR × SR × T × S	18	1.94	1.01
*S*s within groups	60	11.90	6.23*
*S*s within × T	540	1.91	
Total	719	3.49	

Note.—"*S*s within × T" is used as the error term for trials, "*S*s within," and all interactions involving trials. "*S*s within" is used for all other terms.

* $P < .01$.

than a partner's "emotional reaction" to the response. It is possible that the social motive, approval by others, is weak relative to the aroused achievement motive, and *S*s regard the partner's response as essentially "noisy" information. This suggests two further experimental possibilities: eliminating the task reinforcement (indication of correctness) might increase the efficacy of the social reinforcement; and increasing the strength of the motivation for approval might increase the effects attributable to social reinforcement. Two brief exploratory studies were performed to check on these suggestions.

Experiment 1a.—The first point was examined by simply eliminating the task reinforcement. The procedure was identical to that reported above except that the correct answer was not given to *S*s after each trial. At the end of 100 trials they were told how many correct predictions they had made. There were 16 males, run in pairs, 8 under each of the two schedules of social reinforcement, 1.00 for imitation, 1.00 for opposition. Figure 2 indicates that simple elimination of the task reinforcement did not appreciably enhance the effects of the social reinforcement. In general, the level of imitation is higher when imitation is socially reinforced but the difference is not significant.

Experiment 1b.—The second test required some method for strengthening the motivation for approval. This was accomplished by informing *S*s that on the basis of their performance in the present session certain *S*s would be chosen for subsequent studies on decision making for which they would be paid. They were informed that two criteria would be important in selection, their accuracy in prediction and their compatibility. To assess the latter they would be asked to complete a compatibility rating scale on their partner at the end of the session. In all other respects the method was identical to that used in the above described study. There were six males under each of the two schedules of social reinforcement, 1.00 for imitation, 1.00 for opposition.

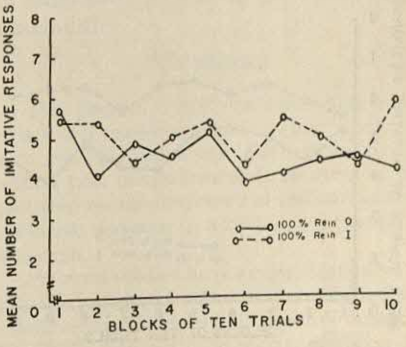


Fig. 2. Mean number of imitative responses per block of trials as a function of probability of social reinforcement: No task feedback.

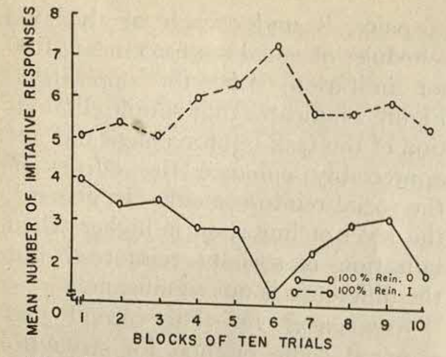


FIG. 3. Mean number of imitative responses per block of trials as a function of probability of social reinforcement: No task feedback; compatibility rating.

Figure 3 indicates that the introduction of the compatibility evaluation did result in a marked effect attributable to the social reinforcement. The *S*s imitated when imitation was approved of by the "partner" and opposed when opposition was approved by the "partner." The mean difference between the conditions is significant ($F = 8.61$, $df = 1/10$, $P < .05$). It appears that *S*s, our college *S*s at least, are not terribly concerned about social approval unless it is instrumental to some more important goal.

The results of a further test of this same possibility, using 20 *S*s per condition, are shown in Fig. 4. The difference between the two conditions is significant ($F = 8.49$, $df = 1/36$, $P < .01$). It should be noted, however, that the effects of a 1.00 schedule of *social reinforcement* appear to be much less than the effects of a .80 schedule of *task reinforcement* (compare .80 conditions in Fig. 1 and 4).

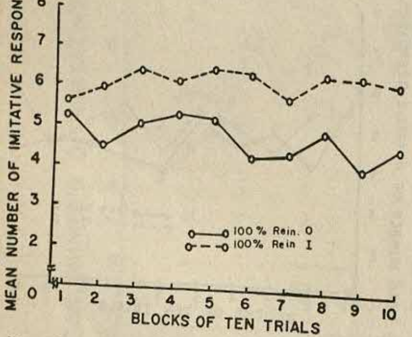


FIG. 4. Mean number of imitative responses per block of trials as a function of probability of social reinforcement: No task feedback; compatibility rating.

DISCUSSION

The results clearly confirm our earlier findings on the effects of different schedules of task reinforcement on the acquisition of an imitative response: the level of imitation is a positive function of the probability of task reinforcement for imitation. However, as in an earlier study using verbal feedback from *E* (Kanareff & Lanzetta, 1960), social feedback did not appreciably affect the probability of an imitative response under either of the task feedback conditions.

The inefficacy of the GSR as a social reinforcer may be attributable to many factors, of which the following appear most plausible:

1. When two cues are available, one of which is specified as relevant to the primary task of performing effectively, *S*s will attend to the cue specified as relevant. Operant conditioning studies, especially those utilizing a verbal reinforcer, typically provide cues which are unspecified; they may be interpreted either as task or social feedback. In the present study *S*s receive task informative feedback as well as independent social feedback which is known to be irrelevant to task performance. Terrell and Kennedy (1957) have demonstrated that when one of two sets of cues is labeled as task informative while the alternative set is left unspecified, the specified set predominates. Our results may be reflecting the operation of such a selective reaction to labeled features of the environment.

2. Subjects may not have been motivated to elicit positive reactions from the

partner. In view of the results for the .5 task reinforcement condition and the data obtained under "no task feedback" conditions the labeling of cues per se postulated above does not appear to be the critical variable. More important may be the cues which define the primary objective of performance. The *S*s may define their task to be that of attaining a high number of correct responses and only cues which are specified as relevant to this task are attended to. When the specified cues are invalid or when no cues are specified as task informative *S*s may attempt to utilize other nonlabeled cues which are available but they will not employ cues which have been specified as nonrelevant by *E*. Implicit in such an analysis is the assumption that the experimental conditions are highly effective in arousing strong achievement motivation; so strong indeed that even when no cues are provided for guiding behavior toward achievement goals *S*s do not respond to cues which allow them to achieve a presumed secondary goal of "pleasing the partner." The results obtained when a more explicit effort is made to arouse or heighten the motivation to "please the partner" lends some support to the assumption that affiliation motives were weak relative to achievement motives in the major study.

3. Subjects did not "perceive" the contingency between the partner's GSR responses and their own behavior but rather assumed that the partner was responding to his own success or failure. This is not to imply that a "perceived" contingency is necessary or sufficient for conditioning; there is some evidence that conditioning can be achieved even when *S*s do not report being aware of a contingency between response and reinforcing cue (e.g., Dailey, 1953; Essman, 1959) although Levin (1961) has recently challenged these findings. Rather it is assumed that a set to associate a cue with an event over which one has no control will attenuate the effectiveness of the cue as a reinforcer.

Unfortunately, no systematic procedure was employed to determine whether *S*s did in fact perceive a relationship between their own responses and partner's GSR. Some *S*s spontaneously volunteered information which indicated that such an association was often made. The same remarks, however, force one to question the assumption that *S*s, if concerned about the partner's reaction at all, are motivated to elicit a positive GSR response. One *S* spoke with glee of his ability to *negatively* arouse his partner.

The two exploratory studies do not efficiently rule out any of these possible explanations of the inefficacy of the social reinforcement when paired with a task reinforcement. Elimination of the task informative feedback did not enhance the effects of the social feedback, which casts some doubt on the viability of Explanation 1, but does not help in discriminating between 2 and 3. Introduction of the additional criteria of "compatibility" did enhance the effects of the social reinforcer but the induction was not evaluated under conditions where task and social feedback are both present. In addition, the increased conditioning with this induction may be attributed to either heightened motivation to please the partner per se (Explanation 2) or to an increase in observational responses of the GSR cue which would increase the likelihood of *S*s perceiving a contingency between their own behavior and the partner's GSR (Explanation 3). Further studies are obviously necessary in order to select among these alternative possibilities.

SUMMARY

The present study examined the effects of a social reinforcement (emotional response of a partner) either congruent or in conflict with a task reinforcement (indication of correctness) on the frequency of utilization of an imitative response in a two-choice prediction situation.

Previous studies have amply demonstrated the efficiency of various social responses as reinforcers for a variety of classes of behavior. Typically, however, the social reinforcer has not been paired with an objective indication of performance adequacy—"social reality" has been the only basis provided for evaluating performance (e.g., Asch, 1956; Festinger,

1954). The present study indicates that for the limited set of conditions employed, at least, a social reinforcer is much less effective in modifying behavior than an objective indication of response adequacy. The efficacy of the emotional response of the partner was enhanced by making it instrumental to other, more remote, goals but even under such conditions it was less effective in modifying response probabilities than a task reinforcement.

REFERENCES

ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

BUSS, A., GERJUOY, I., & ZUSMAN, J. Verbal conditioning and extinction with verbal and non-verbal reinforcers. *J. exp. Psychol.*, 1954, **47**, 106–110.

DAILEY, J. M. Verbal conditioning without awareness. Unpublished doctoral dissertation, State University of Iowa, 1953.

ESSMAN, W. B. "Learning without awareness" of responses to perceptual and verbal stimuli as a function of reinforcement schedule. *Percept. mot. Skills*, 1959, **9**, 15–25.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

FESTINGER, L. Theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68** 409–416.

KANAREFF, V. T., & LANZETTA, J. T. The acquisition of imitative and opposition responses under two conditions of instruction induced set. *J. exp. Psychol.*, 1958, **56**, 516–528.

KANAREFF, V. T., & LANZETTA, J. T. Effects of task definition and probability of reinforcement upon the acquisition and extinction of imitative responses. *J. exp. Psychol.*, 1960, **60**, 340–348.

LANZETTA, J. T., & KANAREFF, V. T. The effects of a monetary reward on the acquisition of an imitative response. *J. abnorm. soc. Psychol.*, 1959, **59**, 120–127.

LEVIN, S. M. The effects of awareness on verbal conditioning. *J. exp. Psychol.*, 1961, **61**, 67–75.

TERRELL, G., Jr., & KENNEDY, W. A. Discrimination learning and transposition in children as a function of the nature of the reward. *J. exp. Psychol.*, 1957, **53**, 257–260.

(Early publication received March 28, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 329–334

RUNNING SPEED IN RATS AS A FUNCTION OF DRIVE LEVEL AND PRESENCE OR ABSENCE OF COMPETING RESPONSE TRIALS[1]

GEORGE A. CICALA[2]

Princeton University

In a recent paper, Estes (1958) has offered a tentative theory of motivation based on the proposition that "drive stimuli are simply stimuli with no special properties whatever" (p. 42). This view contrasts sharply with that of Hull (1951) who attributes to drive stimuli both stimulus properties and special energizing properties.

In Estes' theoretical analysis of drive he argues that increases in running speed as a function of increased drive level are due not to the energizing properties of drives, but rather to the differential sampling probability of drive stimuli and the effects of extraneous stimuli in the learning situation. He maintains that during acquisition both differential sampling of drive stimuli and extraneous stimuli are responsible for differences in running speed usually found at different drive levels. At the asymptote of training he postulates that all drive stimuli are conditioned to the relevant goal response, and that differences in running speed are solely due to the distracting effect of extraneous stimuli.

In an experiment relating to these assumptions as they apply to asymptotic performance, Cotton (1953) compared asymptotic running speed on all trials versus those trials on which extraneous responses were not observed to occur. This technique of analysis was based upon the rationale that although it was impossible to exclude from the experimental situation all extraneous stimuli, this effect might be accomplished by removing from the data all trials on which behavioral evidence indicated the presence of such stimuli. Cotton found that when all trials were included in the analysis, running speed was an increasing function of the period of deprivation. When trials on which competing responses occurred were excluded from the analysis, however, this effect was severely diminished. This finding was interpreted by Estes (1958) as supporting the hypothesis that differences in running speed at different drive levels were attributable to the presence of unconditioned, extraneous stimuli.

EXPERIMENT I

This experiment was designed to study the acquisition of a running response when competing response trials are excluded from the data. The main interest in this type of analysis stems from the point of view that acquisition of an instrumental response may be entirely a function of the elimination of competing responses, although Estes does not suggest this as a specific possibility. A secondary aim of this study was to test Estes' (1958) prediction that learning curves under different levels of deprivation will converge as the asymptote of training is approached when competing response trials are removed from the data.

---

[1] This research represents, in part, a dissertation submitted to the Princeton University faculty in partial fulfillment of the requirements for the PhD degree. The author wishes to express his thanks to Byron A. Campbell for his continuing encouragement and assistance throughout all phases of this work. While this research was being undertaken, the author was a Public Health Service Terminal Year Fellow.

[2] Now Public Health Service Postdoctoral Fellow, Princeton University.

Method

Subjects.—The *S*s used in this study were 48 male albino rats of the Wistar strain, between 90 and 100 days of age at the start of the experiment.

Apparatus.—The apparatus consisted of a 4-ft. alley with a 1-ft. startbox and four removable goalboxes 1 ft. in length. The alley, start, and goalboxes were 6 in. wide. The sides were constructed of aluminum, the floor of stainless steel grids, and the top was of Plexiglas to permit *E* to observe the animal while it was traversing the alley. Running time was recorded on a Standard Electric timer by means of photocells mounted 2 in. from each end of the alley.

Procedure.—Initially, *S*s were placed in individual cages and given ad lib. access to food and water for 3 days. They were then assigned to four groups equal with respect to weight, and placed on a restricted diet of 10, 15, 20, and 25 gm. of Purina laboratory chow which was placed in a food cup in the home cages at 3 P.M. each day. After 11 days of adaptation to this schedule, *S*s were conditioned to run a straight alley for a constant reinforcement of six 45-mg. food pellets manufactured by the P. J. Noyes Company. The *S*s were run between 9 A.M. and 12 N. each day.

The daily procedure was as follows: 4 *S*s were removed from their home cages and placed in the four goalboxes. The *S* was removed from the goalbox and placed in the stationary startbox, after which the goalbox was positioned at the end of the alley. The *S* was permitted 3 min. to enter the alley, after which it was pushed into the start section of the alley and permitted 3 min. to enter the goalbox. If it had not entered the goalbox in the 3 min. alotted, it was placed in the goalbox and left there until all six pellets were consumed. Each *S* of the squad was run once on each rotation until each of the 4 *S*s had received its appointed number of trials for that day.

The *S*s were given 58 trials in the runway. One trial was given on Day 1, two on Day 2, three on Day 3, and four on each subsequent day for 12 days. Running times were measured for each trial, and *E* made a graphic plot of each *S*'s path and competing responses as it went down the alley. Following Cotton's (1953) procedure, a "competing response" was defined as any trial on which *S* stopped, scratched, or exhibited any behavior incompatible with runway traversal.

The experiment was run in four sections. Each part consisted of the following deprivation groups. Part 1: 10 and 20 gm., Part 2: 10 and 20 gm., Part 3: 15 and 25 gm., Part 4: 25 gm. The total number of *S*s at each deprivation level was 12. Since age and environmental conditions were controlled during training, the fact that deprivation levels were not equally distributed over the course of the experiment was considered not an important design deficiency.

In order to establish the reliability of *E*'s judgment concerning *S*'s behavior on each trial, samples of trials in all deprivation groups, as well as for different stages of acquisition, were run with two observers. Out of the total of 125 trials (75 with a trained psychologist as the *O* and 50 with a laboratory technician trained in biology) the observer agreed with *E* as to whether or not a competing response had occurred 119 times (95% agreement), indicating a high degree of accuracy in judging the presence or absence of competing responses. No difference in reliability was found between the untrained technician and the two psychologists.

Results and Discussion

Figure 1 shows the main results of Exp. I. The left-hand portion of the figure shows mean running speed in the alley when both competing and noncompeting response trials are included in the analysis, and the right-hand portion shows running speed when only noncompeting response trials are included in the analysis. At the asymptote of training (Days 10–16) all differences between once-removed groups (e.g., 10 gm. vs. 20 gm.) were significant at beyond the .01 level (the Mann-Whitney *U*, two-tailed test was used for this and all subsequent difference tests) for both analyses. Another important aspect of the data is presented in Fig. 2 which shows the mean number of competing responses observed for each

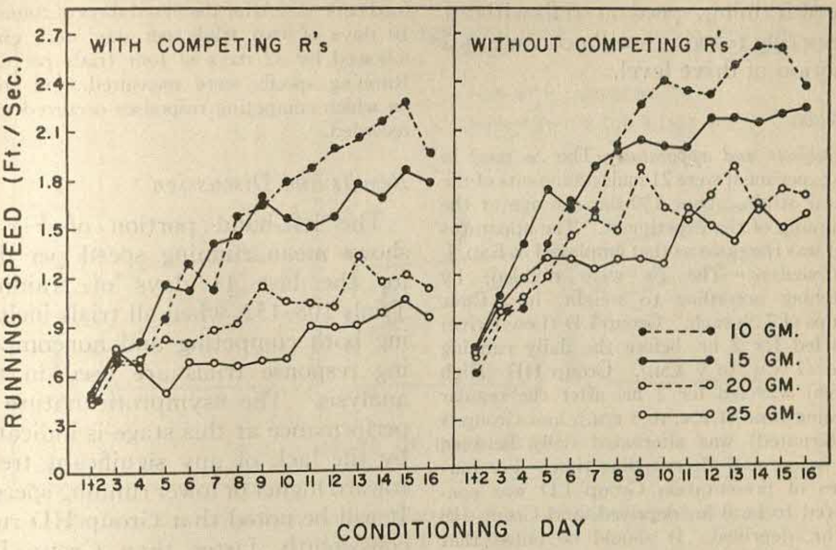


FIG. 1. Mean running speed for all trials (left) and for trials on which no competing response occurred (right): Exp. I.

group for each block of four trials during acquisition. For each point, then, the maximum number of competing responses was 48. Clearly, the number of noncompeting response trials varies with drive level. Differences between once-removed groups were significant at the .02 level or better. In short, the data indicate that the acquisition curve is not entirely determined by the dropping out of competing responses during the early phases of learning. Although the number of competing responses decreases over the course of acquisition, the removal of competing response trials does not serve to decrease the divergence of the acquisition functions for the different drive levels.

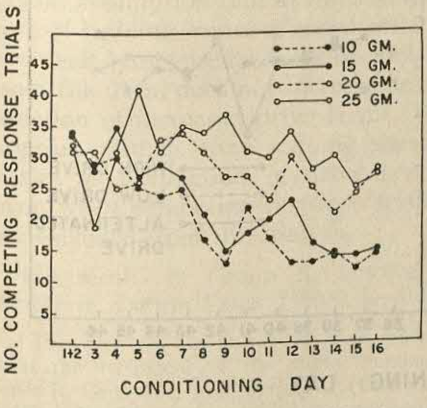


FIG. 2. Mean number of competing response trials over the course of acquisition: Exp. I.

EXPERIMENT II

In Exp. I it was noted that the removal of competing response trials did not cause a greater convergence as the asymptote of training was approached than was found when all the trials were included in the analysis. Because of this finding, Exp. II was designed to study running speed with and without competing response trials as a function of the level of drive at the asymptote of training. Specifically, it was hoped that this study would provide a test of Estes' (1958) assumption that, at the asymptote of training, speed on trials without competing responses will not vary as a function of drive level.

Method

Subjects and apparatus.—The *S*s used in this experiment were 21 male albino rats of the Wistar strain, about 130 days of age at the beginning of the experiment. The apparatus used was the same as that employed in Exp. I.

Procedure.—The *S*s were divided, by matching according to weight, into three groups of 7 *S*s each. Group LD (Low Drive) was fed for 2 hr. before the daily running time (7 A.M. to 9 A.M.). Group HD (High Drive) was fed for 2 hr. after the regular running time (1 P.M. to 3 P.M.), and Group A (Alternated) was alternated daily between the before and after feeding time. For purposes of presentation Group LD was considered to be 0 hr. deprived, and Group HS 18 hr. deprived. It should be noted that using this feeding schedule, all *S*s received roughly the same amount of food each day. This technique resulted in weight changes which were constant among the groups.

Following 16 days of adaptation to this method of feeding, *S*s were given one trial per day in the runway for a period of 4 days. As in Exp. I, each traversal of the runway was regularly reinforced with six 45-mg. food pellets. After the first 4 days of running, 10 days of two trials per day were given followed by 32 days of four trials per day. Running speeds were measured, and trials on which competing responses occurred were recorded.

Results and Discussion

The left-hand portion of Fig. 3 shows mean running speed per day for the last 11 days of training, Trials 108–152, when all trials including both competing and noncompeting response trials are used in the analysis. The asymptotic nature of performance at this stage is indicated by the lack of any significant trend toward higher or lower running speeds. It will be noted that Group HD runs consistently faster than Group LD ($P = .05$). Group A shifts its performance consistently with the daily drive condition ($P = .07$) between the before and after feeding conditions.

When competing response trials are excluded from the data, as shown in the right side of Fig. 3, Group HD

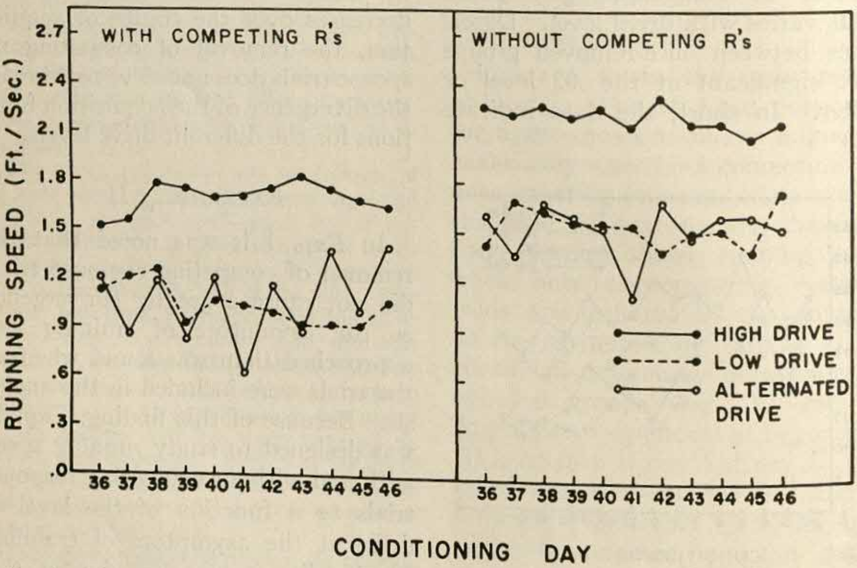


FIG. 3. Mean running speed for all trials (left) and for trials on which no competing response occurred (right): Exp. II. (Group A was fed before running on odd-numbered days and after running on even-numbered days.)

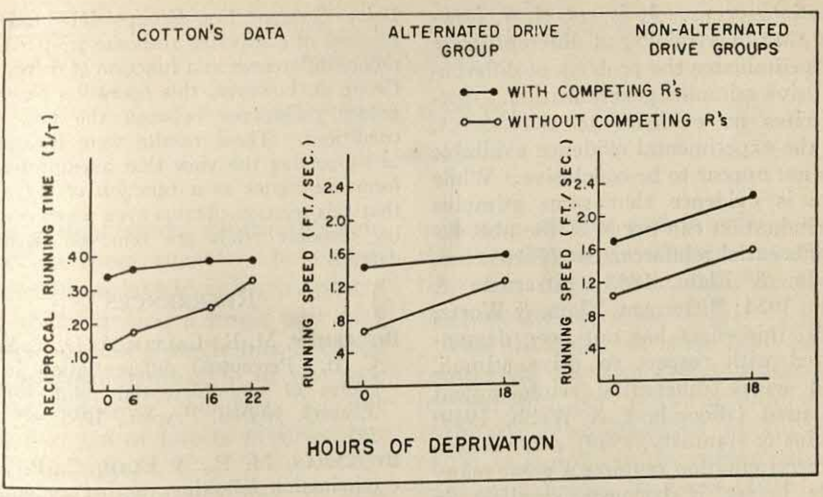


FIG. 4. The comparison between Cotton's data and the results of Groups A, HD, and LD of Exp. II.[3]

remains consistently superior to Group LD ($P = .04$). Again, it should be noted that the functions are asymptotic as indicated by the lack of any significant trend. That the removal of competing response trials fails to diminish the differences in running speed between Groups HD and LD indicates that the differences were not entirely determined by the presence of extraneous cues. This conclusion is clearly contrary to Estes' (1958) assumption that at the asymptote of training, running speed, when competing response trials are excluded from the data, does not increase as a function of increased drive level. It appears, rather, that running speed varies as a function of drive level even when competing response trials are excluded from the analysis.

The results of Group A, however, appear to confirm Estes' (1958) prediction, in agreement with Cotton's (1953) results, since the removal of competing response trials reduces greatly the differences in running speed between the two drive conditions. This relation is shown more clearly in Fig. 4 which compares the results of Group A and the nonalternated groups of the present study with the findings reported by Cotton (1953). For the two graphs plotted from the data of the present study, only the postasymptotic performance (Days 36–46) is included. As can be seen, when the data from Cotton's study are compared with the data from Group A of the present study, the functions obtained are nearly identical. However, when a similar plot is made for Groups HD and LD of the present study, the removal of competing responses does not reduce the slope of the function.

It should be remembered that Cotton (1953) gave all his *S*s extensive training at all drive levels before measuring the asymptotic response of each *S* under each drive level—a procedure very similar to that used with Group A in the present study. This procedure was selected, according to Cotton, in order to eliminate the possible effects of differential drive stimulus generalization by permitting the discrimination of drive stimuli to occur. It is not clear, however, that overtraining at different drive levels eliminates the problem of differential drive stimulus generalization. Cotton cites no evidence to that effect, and the experimental evidence available does not appear to be conclusive. While there is evidence that some stimulus discrimination can occur in the absence of differential reinforcement (Bitterman, Calvin, & Elam, 1953; Bitterman & Elam, 1954; Bitterman, Elam, & Wortz, 1953), this effect has not been demonstrated with respect to drive stimuli. Even where differential reinforcement was used (Bloomberg & Webb, 1949; Jenkins & Hanratty, 1949), drive stimulus discrimination requires a great many trials before it becomes manifest in differential responding.

---

[3] The plot of Cotton's data was made from the reciprocal of the mean running times in Group II, Test Period II (Cotton, 1953). The data were supplied in a personal communication. The graph appears to be the same as that presented by Spence (1956, p. 171).

In the light of these considerations, it may be that overtraining at alternated drive levels does not eliminate the possibility of drive stimulus generalization, and, accordingly, that drive stimulus generalization accounts, in part, for the differences between the alternated and nonalternated groups.

SUMMARY

Two experiments, utilizing Cotton's (1953) technique of analyzing noncompeting response trials separately, were performed to determine the role of extraneous stimuli in the acquisition and performance of an instrumental running response. In Exp. I running speeds were measured during acquisition with amount of food deprivation as the parameter. When competing response trials were removed from the data, normal acquisition functions still obtained and no greater convergence of the drive functions was found than when all trials were included in the analysis. In Exp. II running speeds were measured for a High, Low, and Alternated Drive Group. For Groups HD and LD, removal of competing response trials did not reduce differences as a function of drive. For Group A, however, this operation served to reduce differences between the two drive conditions. These results were interpreted as supporting the view that asymptotic performance varies as a function of drive and that this relation obtains even when competing response trials are removed from the data.

REFERENCES

BITTERMAN, M. E., CALVIN, A. D., & ELAM, C. B. Perceptual differentiation in the course of non-differential reinforcement. *J. comp. physiol. Psychol.*, 1953, 46, 393–397.

BITTERMAN, M. E., & ELAM, C. B. Discrimination following varying amounts of non-differential reinforcement. *Amer. J. Psychol.*, 1954, 67, 133–137.

BITTERMAN, M. E., ELAM, C. B., & WORTZ, E. C. Perceptual differentiation as a function of non-differential reward and punishment. *J. comp. physiol. Psychol.*, 1953, 46, 475–478.

BLOOMBERG, R., & WEBB, W. B. Various degrees within a single drive as cues for spatial response learning in the rat. *J. exp. Psychol.*, 1949, 39, 628–636.

COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, 46, 341–357.

ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958. Pp. 35–69.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

JENKINS, J. J., & HANRATTY, J. A. Drive intensity discrimination in the albino rat. *J. comp. physiol. Psychol.*, 1949, 42, 228–232.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received July 15, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 335–347

A MULTIPROCESS MODEL FOR PAIRED-ASSOCIATE LEARNING[1]

WILLIAM J. McGUIRE[2]

Yale University

The present study synthesizes several recent theoretical trends bearing on paired-associate learning by postulating that learning a single pair may involve, not just a single link between the stimulus member, S, of the pair and its response member, R, but rather a chain of habits involving at least three links.

The first postulated link in this habit chain involves associating with each S a mediating, stimulus-producing response, r. This r is a partial representation of the S to which it becomes connected, encoding some aspect of the S which discriminates that S from the others in the list. Evidence that *S* does tend to encode the S only partially and, more specifically, only with respect to aspects that usefully distinguish it from other Ss in the list, is found in the stimulus predifferentiation studies. Prelearning a set of Rs to a set of Ss is found to result in positive transfer when a new set of Rs are then learned to the Ss provided the Ss are distinguished by the same aspects on both tasks (Goss, 1953). If they are distinguished by a new aspect, however, there is no transfer if the old aspect is missing on the new task (Hake & Ericksen, 1956); and negative transfer results when the old aspect is present and interferes with the encoding of the alternative aspect that has been made significant on the new task (Kurtz, 1955).

The difficulty of this first habit component (S-r) can be manipulated by varying the physical similarity (primary generalization) of the set of Ss, as is done in the present study and, in principle, by Goss (1953), or by varying the number of irrelevant dimensions, as in the typical concept-attainment study (Pishkin, 1960) in which *S* must learn which is the relevant dimension along with the Ss can be discriminated.

The second link in the habit chain involves learning to make the appropriate gross response, R, to each stimulus, s, produced by the mediating, labeling response. That the R of the pair tends to become associated with this mediating s rather than to the S itself has been demonstrated by Bugelski and Scharlock (1952) for experimentally established mediators and by Russell and Storms (1955) for those provided by language habits. That the mediators tend to be the discriminating labels attached to the Ss is demonstrated by McAllister's (1953) finding that stimulus predifferentiation results in positive transfer to the extent that the differentiating labels (r) are "relevant" to the new Rs to be learned for the Ss.

The third link in the habit chain which we postulate involves the possibility that the response member of the pair, R, may itself be a chain of habits that have to be learned. The importance of this "response learning" has been reviewed by Mandler (1954) and most recently demonstrated by the response-meaningfulness (m') studies (Hunt, 1959; Nobel & McNeely, 1957) and the response-similarity studies (Feldman & Underwood, 1957; Underwood, Runquist, & Schulz, 1959) which deal with two somewhat different aspects of R learning—within-R synthesis, and between-R discrimination.

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements of the PhD degree at Yale University in 1954. The author is greatly indebted to Carl I. Hovland for his advice and guidance and also to C. E. Buxton, R. P. Abelson, and N. E. Miller.

[2] Now at the Department of Social Psychology, Columbia University.

In summary, learning each pair in a paired-associate task involves, not a single S-R connection, but rather three component habits. The first involves S and r; the second, s and R; and the third (which may be a chain of habits) the R elements, R_a, R_b, $\cdots R_n$. Hence, learning a pair may involve the following habits:

$$S \rightarrow r \rightarrow s \rightarrow R_a \rightarrow s_a \rightarrow R_b \rightarrow s_b \rightarrow \cdots \rightarrow R_n$$

We are ignoring here the remote associations that probably get formed and which may account for some slight discrepancies from the predictions reported below.

The materials used in the present study were designed to enable us to identify in each erroneous instance (that is, a failure by the *S* on any trial to respond to a given S with the correct R) which of the links in the chain failed in that instance. By such an analysis of the errors we can determine the practice functions, not only of the usual H_c (the proportion of pairs correctly anticipated on a given trial), but also of H_1 (the proportion of correct S discriminations, i.e., S-r habits, on a given trial), of H_2 (the proportion of correct associations, i.e., s-R_a habits), and of H_3 (the proportion of Rs correctly synthesized, i.e., $R_a \rightarrow s_a \rightarrow \cdots \rightarrow R_n$ habits).

Besides indicating processes involved and sources of difficulty in paired-associate learning, these separate indices enable us to compute independently the curves for stimulus generalization and for intrusion errors, the confusion between which has given rise to that hardy perennial among verbal learning controversies, the temporal trend of stimulus generalization during practice, which has recently blossomed again among Murdock (1958, 1959), Battig (1959), Gibson (1959), and Runquist (1959).

METHOD

Materials and experimental variations.—The Ss and Rs in this study were designed to allow independent manipulation of the difficulty of each of the three sets of habits, H_1, H_2, and H_3.

Each *S* learned, by the anticipation method, a list of nine paired associates. The Ss of these pairs were nine, solid black circles of varying diameters. Two different sets of circles were used for different groups of *S*s. In Set I the diameters ranged from .37 to 1.49 cm. in .14-cm. steps; in Set II, from .37 to .93 in .07-cm. steps. The difficulty of learning the S-discrimination habits (H_1) was defined as being greater for Set II than for Set I.

The Rs were numbers. For some *S*s these Rs were the numbers from 1 to 9. For other *S*s the Rs were 3-digit numbers, each beginning with a different integer between 1 and 9, and with the second and third digits assigned by a random procedure. The difficulty of learning the R-chaining habits (H_3) was defined as being greater for the 3- than the 1-digit Rs. In fact, for the 1-digit numbers, H_3 was defined as having a value of 1.00 from the outset of the experiment; that is, it was assumed that *S* would always perform perfectly at this segment of the task.

The difficulty of learning the s-R_a association habits (H_2) was manipulated by the manner of assigning Rs to Ss. An advantage of using sets of Ss that differ along a single linear dimension, size in this case, is that it permits hypothesizing what the discriminating labeling rs will be. With the nine different-sized but otherwise similar Ss used here and with the instructions to *S* described below, it was assumed that *S* would use a numerical-type labeling response; for example, he would tend to label the smallest circles as "one" the next larger as "two," etc., up to the largest which he would label as "nine." For some *S*s the Rs were assigned consecutively so that the successively larger numbers were assigned

TABLE 1

DESCRIPTION OF DIFFICULTY LEVELS OF COMPONENT HABITS IN THE SIX CONDITIONS, WITH FORMULAS FOR PROPORTIONS OF CORRECT ANTICIPATIONS (H_c) AND OF INTRUSION ERRORS (IE)

Cond.	Task			H_c[a]	IE[a]
	H_1	H_2	H_3		
$H_1H_2H_3$	easy	easy	easy	H_1	$1 - H_1$
$H_1'H_2H_3$	hard	easy	easy	H_1'	$1 - H_1'$
$H_1H_2H_3'$	easy	easy	hard	$H_1 \cdot H_3'$	$H_3'(1 - H_1)$
$H_1'H_2H_3'$	hard	easy	hard	$H_1' \cdot H_3'$	$H_3'(1 - H_1')$
$H_1'H_2'H_3$	hard	hard	easy	$H_1' \cdot H_2'$	$1 - H_1' \cdot H_2'$
$H_1'H_2'H_3'$	hard	hard	hard	$H_1' \cdot H_2' \cdot H_3'$	$H_3'(1 - H_1' \cdot H_2')$

[a] These H_c and IE formulas are based on the general equations $H_c = H_1 \cdot H_2 \cdot H_3$ and $IE = H_3(1 - H_1 \cdot H_2)$ and the assumptions that the values of H_2 and H_3, without primes, equal 1.00 from the beginning of practice.

to progressively larger circles. For other groups, Rs were assigned at random to Ss so that no simple system related size of numbers and size of circles. The difficulty of learning the association habits (H_2) was, of course, defined as being greater with the latter, random method of assignment. In fact, in the conditions with consecutive assignment of the numerical responses, H_2 was defined as equaling 1.00 from the outset of practice; that is, it was assumed that *S* always performed this segment of the task perfectly.

Experimental design.—An incomplete 2^3 factorial design (minus one quadrant) was employed. The three variables were difficulty of learning the H_1, H_2, and H_3 sets of habits, there being an easy and difficulty level of each as defined above. Table 1 shows the levels of each of these factors in the six conditions used, with primes (e.g., H_1') indicating the difficult condition and no primes (e.g., H_1) indicating the easy condition on the given link. Ten *S*s served in each of the six conditions, and each *S* served in only one condition.

Apparatus.—A Hull memory drum with 1-in.-square windows was used. The Ss were presented in the left window, and the Rs in the right window. The anticipatory method was employed, S appearing alone for 2 sec. and then S and R appearing together for 2 sec. after which both windows closed again. Almost immediately, the left window reopened, showing the next S on the list. The *S* was required to anticipate the R within the first 2-sec. period. The pairs came in a different order on each trial.

Instructions to S.—The *S*s were given the usual instructions for a paired-associate task regarding how the materials would be presented and what they were to try to do. In addition, they were told that all Ss were solid black circles, differing only in size, and that there were nine different pairs in all. Further, those in H_3 groups were told that the Rs would be the 1-digit numbers from 1 to 9, while those in H_3' groups were told that the Rs would be nine 3-digit numbers, each beginning with a different integer from 1 to 9. Subjects in H_2 conditions were given the additional information that progressively higher numbers were assigned to successively larger circles, while those in H_2' conditions were told that the magnitudes of the Rs of the pairs were randomly related to the sizes of the circles. All *S*s were instructed to try to give some response to each S after the completion of the first trial. Those in H_3' (3-digit R) conditions were further directed to respond with at least the first digit even if they could not give the other two. As a further measure to keep omissions at a minimum, all *S*s were told that when they were not sure of a correct R they should guess.

Length of the learning session.—The experimental session for each *S* consisted of 80 continuous trials without any intertrial pause (48 min.). Only 1 *S* asked to be relieved before completing 80 trials; he was replaced by the next *S* in the pool.

Subjects.—The 60 *S*s were selected on the basis of availability from a pool of several hundred college students in an introductory psychology course and assigned to one of the six conditions in accordance with a random order. The writer served as *E* for all *S*s.

RESULTS AND DISCUSSION

Correct anticipations (H_c): Group means.—The manipulation of the difficulty of each of the three postulated sets of habit links (H_1, H_2, and H_3) did produce the predicted difference in the mean proportion of correct anticipations per trial (H_c) over the 80 trials.[3]

The difficulty of forming the first set of habit links (S-r), involving S discrimination, was manipulated by using .14-cm. steps between the Ss in some conditions (H_1-) and .07-cm. steps in other conditions $(H_1'-)$. Insofar as this set of habits is involved in learning the paired associates, performance (H_c) in Cond. $H_1H_2H_3$ should have been superior to that in Cond. $H_1'H_2H_3$ since the only difference between the two tasks was that the S discrimination was more difficult in the latter conditions. For like reasons, performance on Cond. $H_1H_2H_3'$ should have been superior to that in Cond. $H_1'H_2H_3'$. Both of these predictions are confirmed by the results. The mean H_c value for the 10 *S*s over all 79 anticipatory trials is significantly higher (see Table 2) in Cond. $H_1H_2H_3$ than in $H_1'H_2H_3$ ($P < .001$) and in Cond. $H_1H_2H_3'$ than in $H_1'H_2H_3'$ ($P = .02$).

TABLE 2

PROPORTION OF CORRECT ANTICIPATION PER TRIAL (H_c) IN EACH OF THE SIX CONDITIONS

Cond.	Mean	*SD*
$H_1H_2H_3$	.793	.048
$H_1'H_2H_3$	.635	.057
$H_1H_2H_3'$	.618	.090
$H_1'H_2H_3'$	.491	.116
$H_1'H_2'H_3$	.462	.105
$H_1'H_2'H_3'$	.340	.098

Note.—Each mean is based on 10 *S*s over 80 trials, 9 pairs per trial.

The difficulty of learning the habit chains of R elements was varied by using 1-digit Rs in some conditions $(-H_3)$ and 3-digit Rs in others $(-H_3')$. Hence, higher H_c scores are predicted in Cond. $H_1H_2H_3$ than in $H_1H_2H_3'$; in Cond. $H_1'H_2H_3$ than in $H_1'H_2H_3'$; and in $H_1'H_2'H_3$ than in $H_1'H_2'H_3'$. All the obtained H_c differences (see Table 1) are in the predicted direction and the three differences are significant at the .001, .001, and .01 levels, respectively.

The difficulty of forming the third set of habit links, the s-R_a or "association" habits, was manipulated by assigning Rs to Ss consecutively in some conditions $(-H_2-)$ and randomly in others $(-H_2'-)$. Hence, the H_c scores should be higher in Cond. $H_1'H_2H_3$ than $H_1'H_2'H_3$ and in Cond. $H_1'H_2H_3'$ than $H_1'H_2'H_3'$. Both mean differences are in the predicted direction (see Table 2) and are significant at the .001 and .01 levels, respectively.

So far only the directions of the differences in H_c scores between the conditions have been considered, but the relative sizes of these differences are also of interest. In Table 1 the equations of H_c in each of the six conditions are given. Solving for H_3' in each of the equations where it appears, it is found that in Cond. $H_1H_2H_3'$, $H_3' = H_c \div H_1$; in Cond. $H_1'H_2H_3'$, $H_3' = H_c \div H_1'$; and in

---

[3] The H_c symbol has been deliberately selected in the present model for its resemblance to the Hullian habit strength symbol, $_sH_R$, because the constructs are similar in that both are negatively accelerated increasing functions of the number of reinforced S-R pairings. However, the H_c differs quantitatively from the Hullian $_sH_R$ in three respects. First, the present H_c is a direct one-to-one function of the probability of correct response, whereas $_sH_R$ is an intervening construct that is an exponential function of this probability. Secondly, $_sH_R$ is the strength of a single S-R habit, while H_c is the mean strength of a set of as many habits as there are pairs. Thirdly, the present H_c need not be zero when $N = 0$ since, unlike Hull (1943, 1952), we allow the a and c parameters in the equation $H_c = c - ae^{-bN}$ to assume different values (Lewis, 1960).

Cond. $H_1'H_2'H_3'$, $H_3' = H_c \div (H_1' \cdot H_2')$, the value of H_c in each case being that for the given condition. But the divisors of the right hand members of each of these three equations can be seen in Table 1 to be the H_c values of, respectively, Cond. $H_1H_2H_3$, $H_1'H_2H_3$, and $H_1'H_2'H_3$. Since each of the three fractions is equal to the same thing, H_3', then the following equation can be derived:

$$H_3' = \frac{H_c \text{ of Cond. } H_1H_2H_3'}{H_c \text{ of Cond. } H_1H_2H_3} = \frac{H_c \text{ of Cond. } H_1'H_2H_3'}{H_c \text{ of Cond. } H_1'H_2H_3} = \frac{H_c \text{ of Cond. } H_1'H_2'H_3'}{H_c \text{ of Cond. } H_1'H_2'H_3}$$

Substituting the obtained mean H_c values (shown in Table 2) in this equation, the values of these three fractions are found to be .779, .773, and .736, respectively.

We would expect these three quotients to be equal to the extent that the learning of any one of the three postulated sets of habits proceeded independently of the difficulty of the other two sets of habits being acquired simultaneously. The closeness of the three obtained quotients suggests that there is considerable independence. The closeness of the .779 and .773 values, particularly, suggests that the difficulty of learning the response chains is only negligibly affected by the difficulty of the S discriminations being learned concurrently. The third value, .736, is slightly (about 5%) lower than the other two. While the first two values give a derived measure of H_3' in consecutive S-R assignment conditions, this third gives an H_3' value in a condition where the numerical values of Rs are randomly related to the size of Ss to which they are assigned. Hence, the learning of the response chain may be (very slightly) impeded by the difficulty of the s-R association habits being learned concurrently. This interaction might be expected in view of the similarity of the s-R habits and the $R_a \rightarrow s_a \rightarrow \cdots R_n$ habits in the present study (both involve linking numbers) and of the usual finding that difficulty of learning is a positively accelerated function of amount of material (Hovland, 1940).

These derived estimates of H_3' in Cond. $H_1H_2H_3'$ and Cond. $H_1'H_2H_3'$ can be checked by an independent method of computing H_3' which requires us to anticipate the classification of errors discussed below. Briefly, the sum of Categories A and C in Table 3 should yield a running measure of H_3', the number of R chains correctly learned. The mean H_3' for Cond. $H_1'H_2H_3'$ computed in this way is .778 (as compared to the .773 mean yielded by the quotient method just discussed). The mean for Cond. $H_1H_2H_3'$ is .800 (as compared with the .779 mean yielded by the quotient method). The slight discrepancy in each condition, between the summation and the quotient method of deriving H_3', falls far short of conventional significance levels.

The H_c learning curves.—We can obtain a more analytic understanding of the effect of manipulating H_1, H_2, and H_3 on proportion of correct anticipation (H_c) if the group trends during practice in each of the six conditions are considered. To calculate these practice functions, the group mean H_c values in each condition over the 80 trials were used to determine 13 practice points, as follows. Trial 1 was omitted in the determination of the fitted curves, since this trial constituted an initial presentation to allow *S* to see each pair at least once before being required to anticipate. The first point (point "2") is based on scores on Trials 2 and 3, the next point (point "5") is based on Trials 4, 5, and 6; the next (point "8") on Trials 7, 8, 9, and 10; and each of the next 10 points (called 14, 21, 28, 35, 42, 49, 56, 63, 70, and 77, respectively) on successive blocks of seven trials. For example, point 77 is based on the means of Trials 74 through 80. More points were calculated in the case of the earlier trials in order to determine more accurately the shape of the function near the beginning of practice, when improvement was most rapid. The scores at each practice point for each of the six groups are plotted in Fig. 1, which also shows the fitted function for each condition.

All of the functions shown in Fig. 1 are of the inverse exponential family; $H_c = c\text{-}ae^{-bN}$, where H_c is the proportion of the nine possible anticipations given correctly on any trial, c-a is the proportion of correct anticipations at the outset of practice, c is the asymptote, and b, the growth parameter. Exponential functions were used because they described the data more accurately than any straight line in each of the six conditions by an amount that was significant at the .05 level in all cases except Cond. $H_1'H_2H_3$. In Cond. $H_1H_2H_3$ and $H_1'H_2H_3$ we also fitted hyperbolic and Gompertz functions to the data, but these functions were found to describe the data somewhat less adequately than the exponential. Cond. $H_1H_2H_3'$ was the only one of the six in which the data deviated from the fitted exponential function by an amount that approached the .05 level. The test for goodness of the fit was based on Lindquist's Case 8 (1947) with $df = 10$ (i.e., $13 - 3$, the number of trial blocks minus the number of parameters calculated for each function).

Some criticism has been published by Sidman (1952) and Bakan (1954) regarding the practice of inferring the shape of the individual functions from that of the group functions, especially in the case of exponential functions. The criticism has no bearing on the results reported in the

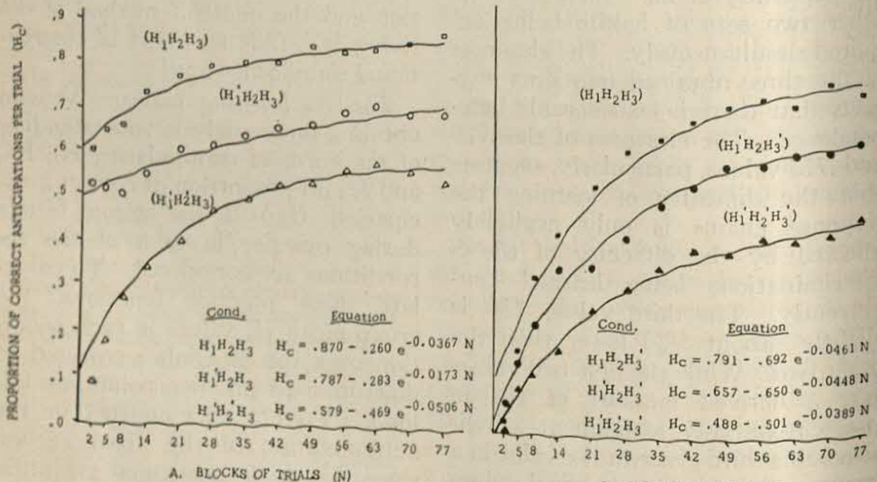


FIG. 1. Mean proportion of correct anticipations (H_c) on each block of trials. (Figure 1a shows curves for one-digit, and Fig. 1b, for three-digit response conditions.)

present experiment since no inference is made regarding the exact shape of the individual curves. The objection is relevant at all, only when there is an appreciable variance among the growth parameters (b) of the individual curves, and when the $b \cdot N$ products are large. The second condition does obtain in the present study, but whether there is a significant variance among the individual b values has not been computed (because of labor involved in fitting exponential curves to the data of each of the 60 individuals). In the absence of more information on the b parameters, caution should be exercised in inferring the shapes of the individual *S* curves in the present study.

Effect of the experimental variables on the different parameters of the H_c learning curves.—Several conclusions can be drawn from the functions obtained in the present study. Each of the three hypothesized sets of habits is found to play a consistent part in determining both the initial level of performance (the H_c-intercept, equal to c-a) and the asymptote (c) of the functions. This conclusion follows from the obtained effect of increasing the difficulty of either the H_1, H_2, or H_3 task, which is found in all cases to lower both of these parameters. There are seven nonredundant pairs of conditions between which such a comparison is possible, and for both the c and the c-a parameters, the above stated effect was obtained in all seven comparisons.

The growth parameter (b) of H_c appears to be related to the difficulty of the component sets of habits, H_1, H_2, and H_3, in a more complex way. In the first place, it seems that where the source of greater difficulty derives from the greater similarity between the *S*s, so that there are more generalization errors, then the rate of learning of the paired-associate task is slower as the difficulty increases (compare b parameters in Cond. $H_1H_2H_3$ and $H_1'H_2H_3$ and in Cond. $H_1H_2H_3'$ and $H_1'H_2H_3'$). As regards the effect on the growth parameter of increasing the difficulty of the R-chains (H_3) or of the s-R associations (H_2), the b parameters shown in Fig. 1 suggest that increasing the difficulty of one or both of these increases the rate of learning (compare Cond. $H_1H_2H_3$ with $H_1H_2H_3'$; and Cond. $H_1'H_2H_3$ with $H_1'H_2H_3'$, with $H_1'H_2'H_3$, and with $H_1'H_2'H_3'$), but that increasing the difficulty of just one (H_2 or H_3) increases the learning rate more than does increasing the difficulty of both H_2 and H_3 (compare Cond. $H_1'H_2H_3'$ with $H_1'H_2'H_3'$; and Cond. $H_1'H_2'H_3$ with $H_1'H_2'H_3'$). These interpretations are post factum, however, and call for further investigation.

Practice curves of stimulus discrimination and of intrusion errors.—An intrusion error is defined as the giving of an R that is in the list but giving it to an S other than the one with which it is actually paired in the list as presented. Hence, the probability of an intrusion (IE) is directly proportional to the learning of the R chains (H_3) and inversely proportional to the learning of the S discrimination (S-r) habits (H_1) and of the association (s-R_a) habits (H_2); that is:

$$IE = H_3(1 - H_1 \cdot H_2)$$

An S-generalization error, on the other hand, was defined as giving the wrong r to an S and measured by $1 - H_1$, a value that will tend to differ from IE when either H_2, H_3, or both are less than 1.00. The practice of defining stimulus generalization during paired-associate learning has, as was pointed out above, led to considerable controversy. An *E* may be allowed a certain liberty in defining his terms, as long as he does so clearly. However, this definition that S discrimination has occurred only when *S* discriminates the Ss with the Rs that *E* has arbitrarily assigned becomes misleading, particularly as regards the practice curve of the generalization function and particularly when the Rs themselves or their relations to the Ss must be learned (that is, when H_3 or H_2 increases with practice). Eleanor Gibson (1940, 1942), whose provocative research using the IE definition of S generalization has given rise to this controversy, has herself pointed out (Gibson, 1959; Gibson & Gibson, 1955) the inadequacy of this definition.

The materials used in this study were designed to allow the responses given by *S*s in each instance to be classified into separate categories in terms of which the practice curves of H_1, H_2, and H_3, as well as of H_c, could be calculated. In this way, we could test the above formulations. Each of *S*'s 711 responses (9 pairs × 79 trials) was put in one of the following five categories:

A. A first digit appropriate to the given S (indicating a correct discrimination) and the second and third digits appropriate to the first (indicating a correct R). The proportion of responses in this category equals, by definition, H_c, the usual measure of a correct response in paired-associate learning.

B. A first digit appropriate to the given S (indicating a correct discrimination) and second or third digits not appropriate to the first (indicating a wrong R).

C. A first digit not appropriate to the given S (indicating a wrong discrimination or a wrong s-R association or both) and second and third digits appropriate to the first (indicating a correct R). The proportion of responses that falls in this category equals, by definition, IE.

D. A first digit not appropriate to the given S (indicating a wrong S discrimination or a wrong s-R association or both) and second or third digit not appropriate to the first (indicating a wrong R).

E. An omission, no response given.

The proportions of responses falling into each of these five categories for Cond. $H_1H_2H_3'$ and $H_1'H_2H_3'$ during successive practice intervals are shown in Fig. 2 and Table 3. Only for those two conditions are all categories used unambiguously. In Cond. $H_1H_2H_3$, $H_1'H_2H_3$, and $H_1'H_2'H_3$ where initial $H_3 = 1.00$, i.e., the Rs are one digit numbers, no Category B or D responses would occur. In Cond. $H_1'H_2'H_3$ and $H_1'H_2'H_3'$ (where Rs are randomly assigned to Ss), a wrong first digit indicates that either the S discrimination, or the s-R association, or both, are wrong. Hence, in these two conditions the interpretation of Category C and D responses is ambiguous, and separate measures of H_1' and H_2' are not possible.

Representing the proportion of responses that fall into these categories by the letters A, B, C, D, and E, respectively, then in all six conditions, the pro-

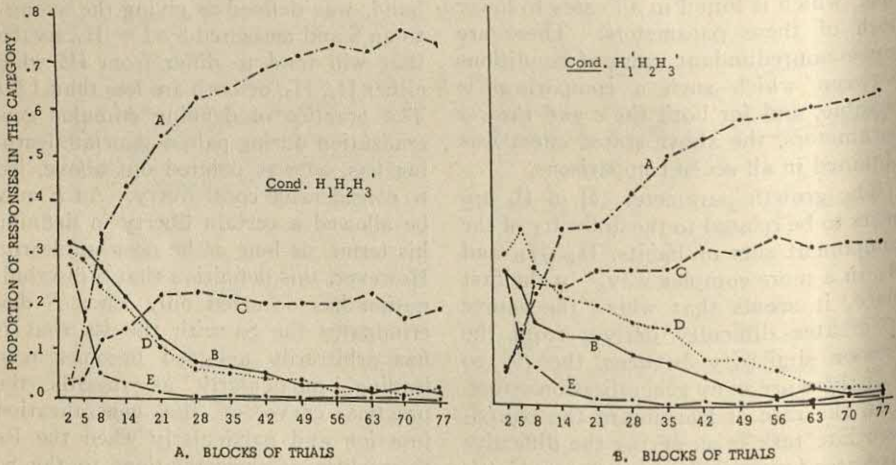


FIG. 2. Practice trends of the mean proportion of responses in each of the five categories (as discussed in text) for Cond. $H_1H_2H_3$ (Fig. 2a) and Cond. $H_1'H_2H_3'$ (Fig. 2b).

TABLE 3

Proportion of the Obtained Responses That Fell in the Five Response Categories in Each Condition

Category	Conditions					
	$H_1H_2H_3$	$H_1'H_2H_3$	$H_1H_2H_3'$	$H_1'H_2H_3'$	$H_1'H_2'H_3$	$H_1'H_2'H_3'$
A. Correct Response	.793	.635	.618	.491	.462	.340
B. Correct S-r and s-R but inadequate R chain	—[a]	—	.094	.079	—	.020
C. Incorrect S-r or s-R, but adequate R chain (=IE)	.207	.365	.182	.287	.436	.330
D. Incorrect S-r, and inadequate R chain	—	—	.083	.118	—	.118
E. Omission	.000	.000	.023	.025	.002	.192
Total	1.000	1.000	1.000	1.000	1.000	1.000

Note.—Each mean is based on 10 *Ss* over 80 trials.
[a] Category B and D responses cannot occur in the three 1-digit R conditions.

portion of correctly learned R chains, H_3, equals A and C and the proportion of pairs that have been mastered to the extent of both correct S-r discriminations and s-R associations, $H_1 \cdot H_2$, equals A and B. Since, as discussed above, the proportion of intrusion errors, IE = $H_3(1 - H_1H_2)$, we find by substitution that IE = $(A + C)(1 - A - B)$.[4]

With these preliminaries understood it is possible to test the hypotheses concerning stimulus discrimination and intrusion errors as a function of practice. In Cond. $H_1H_2H_3$ and $H_1'H_2H_3$ (the two conditions with consecutively-assigned one-digit Rs) the test is a direct one since initial H_2 and $H_3 = 1.00$ and there are not omissions. Hence, $H_c = H_1 \cdot H_2 \cdot H_3$ becomes $H_c = H_1$, i.e., in these two conditions the overall performance score on the paired associates (H_c) is identical with that of stimulus discrimination (H_1). In both of these conditions there is a clear tendency for H_1 to increase monotonically throughout practice. The trends in both conditions are well fitted by rising, negatively accelerated, exponential functions (see Fig. 3). These results tend to confirm the hypothesis that discrimination increases monotonically during practice with a set of paired associates and to infirm the opposing hypothesis that this function first rises, then falls.

Since initial H_2 and $H_3 = 1.00$ and no omissions occur in these two conditions ($H_1H_2H_3$ and $H_1'H_2H_3$), the equation, IE = $H_3(1.00 - H_1 \cdot H_2)$, simplifies to IE = $1.00 - H_1$, or its equivalent, IE = $1.00 - H_c$. This congruency between the stimulus generalization curve and the intrusion error curve is a special case that occurs, as described by the

---

[4] These equations apply in conditions where there are no omissions or where it is assumed that such omissions as do occur represent responses which, had they occurred, would all have been Category D responses, that is, completely wrong. In the present study a different assumption has been made, namely, that omissions represent responses that, had they been made, would have divided among the other error categories (B, C, and D) in the same proportion as the erroneous responses that were made. Hence, the B and C values used in the subsequent discussion are "corrected" values (represented by the symbols B′ and C′), equal to the obtained B and C values, plus a proportional amount of the E values, yielding the "corrected" equations of $H_1 \cdot H_2 = A + B'$ and IE = $(A + C')(1 - A - B')$. There are still other assumptions regarding omissions that might have been made. This discussion of omissions is important mainly in connection with Cond. $H_1'H_2'H_3'$, since the number of omissions was negligible in Cond. $H_1H_2H_3$ and $H_1'H_2H_3$ and occurred in appreciable numbers in the other three conditions, only during the first few trial intervals.

TABLE 4

Proportion of Correct S Discriminations per Trial (H_1) in the First Four Conditions

Cond.	Mean	*SD*
$H_1H_2H_3$	.793	.048
$H_1'H_2H_3$	.635	.057
$H_1H_2H_3'$	.712	.101
$H_1'H_2H_3'$	.570	.123

Note.—Each mean is based on 10 *S*s over 80 trials, pairs per trial.

TABLE 5

Proportion of Intrusion Errors (IE) per Trial in Each Condition

Cond.	Predicted Mean	Obtained Mean	Obtained *SD*
$H_1H_2H_3$	.207[a]	.207	.048
$H_1'H_2H_3$	.365[a]	.365	.057
$H_1H_2H_3'$	.199	.182	.058
$H_1'H_2H_3'$	.305	.287	.114
$H_1'H_2'H_3$	.436[a]	.436	.109
$H_1'H_2'H_3'$	.348	.330	.092

Note.—Each obtained score is based on 10 *S*s over 80 trials, 9 pairs per trial. Predicted scores were derived from the equation: $IE = (A + C')(1 - A - B')$.

[a] In the three 1-digit R conditions, the "predicted" means necessarily coincide with the obtained means.

model, when initial H_2 and $H_3 = 1.00$ and there are no omissions. When the IE curve does coincide with that of stimulus generalization, it, as well as the latter, falls throughout practice.

In the two conditions with consecutively assigned 3-digit Rs, $H_1H_2H_3'$ and $H_1'H_2H_3'$, the equation $H_1H_2 = A + B'$ reduces to $H_1 = A + B'$ on the assumption that $H_2 = 1.00$ in the consecutive assignment conditions. (This assumption may be somewhat extreme, in which case the values of H_1 shown in Table 4 and Fig. 3 would slightly underestimate the true values.) Figure 3 shows the practice curves for S discrimination in these two conditions, which can be seen to follow the predicted monotonically rising trend, well fitted by negatively-accelerated exponential functions, rather than the sometimes reported rising, then falling curve.

The intrusion error values (Table 5) are given by the formula $IE = (A + C')(1 - A - B')$. These predicted practice functions for IE in Cond. $H_1H_2H_3'$ and $H_1'H_2H_3'$ are shown in Fig. 4, where the obtained values (that is, number of category C responses) are also shown.

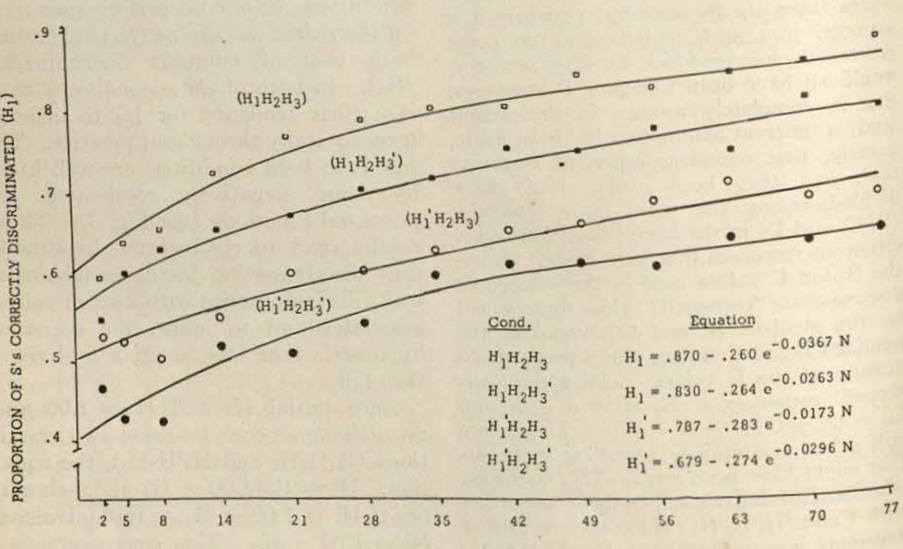


Fig. 3. Mean proportion of correct S discriminations (H_1) on each block of trials for the four conditions in which this information was available.

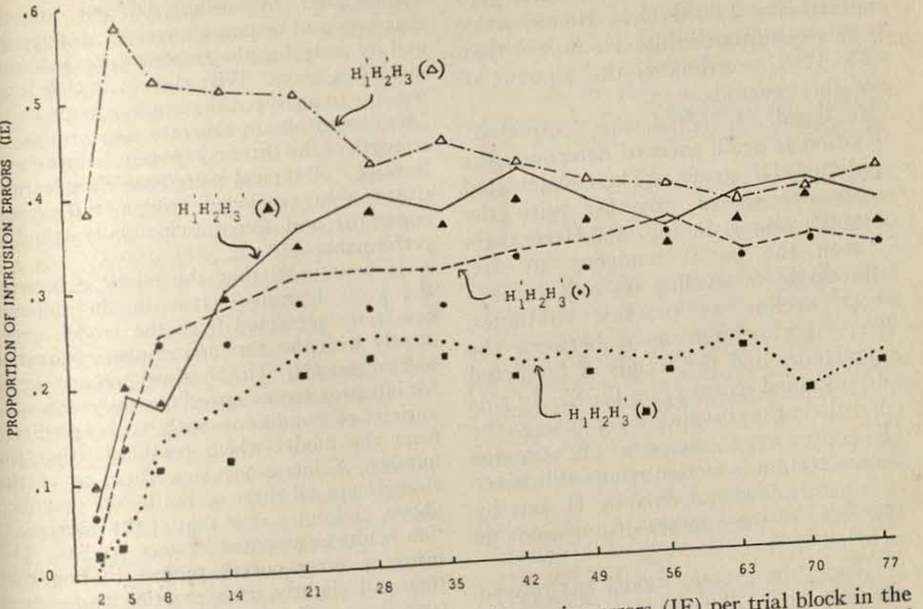


FIG. 4. Practice trends for mean proportion of intrusion errors (IE) per trial block in the four conditions in which Category B or E responses occurred. (Lines indicate predicted trends; symbols, the obtained values in the given condition.)

It can be seen that in each condition there is good agreement between predicted and obtained scores, with respect to practice trend and maximum point, as well as to height. There is a slight tendency in both conditions for the predicted values of IE to overestimate the obtained values but these discrepancies yield *F* values of only 1.14 and 0.61 in the two conditions (using the *S*s × Trials interaction as the error term), and therefore such discrepancies as these are can be attributed to chance. These IE curves in both conditions show an initial rise, followed by a leveling off and a slow decline, while the stimulus generalization trend has just been seen to follow a falling course throughout practice. This disagreement demonstrates that, while the shape of the IE function does coincide with that previously reported (Gagné, 1950; Gibson, 1942; Underwood & Goad, 1951) for S generalization, it does not coincide with S generalization more logically defined.

For the final two conditions, $H_1'H_2'H_3$ and $H_1'H_2'H_3'$, with randomly paired Ss and Rs, the response categorization possible with the present materials yields separate values of H_3 and of $H_1 \cdot H_2$, but not of H_1 and H_2 separately, as discussed above. Hence, it is not possible to test the hypotheses regarding S discrimination with the results from these conditions. The hypotheses regarding IE, however, can be tested since for this a composite value for $H_1 \cdot H_2$ suffices.

Since in Cond. $H_1'H_2'H_3$, initial $H_3 = 1.00$, the predicted $IE = 1.00 - A - E$. The values for IE so calculated necessarily coincide with the obtained values since there are no category B or D responses in this condition. The practice trend shows a brief initial rise followed by a subsequent decline (see Fig. 4). Here again it can be seen that intrusion errors do not furnish a valid measure for stimulus generalization: first, because it is influenced also by omissions (which account for the brief initial rise); and secondly, because aside from the omissions which become infrequent after the first few trial intervals, the

$IE = 1.00 - H_1 \cdot H_2$ while stimulus generalization $= 1.00 - H_1$. Hence, since in this condition initial H_2 is less than 1.00, IE overestimates the amount of stimulus generalization.

In Cond. $H_1'H_2'H_3'$, the "corrected" equation is again used to determine the predicted IE scores. These calculated scores are shown, together with the obtained scores, in Fig. 4. Here again is seen the initial tendency to rise, followed by a leveling off and perhaps slight decline as practice continues. There is close agreement between the magnitudes and the trends of predicted and obtained scores ($F = 0.70$). In this condition the invalidity of using the IE score as a measure of stimulus generalization is more obvious still, since it is determined not only by H_1 but by H_2, H_3, and the number of omissions as well.

Hence, the present model has proved adequate to predict the IE curves over a wide range of conditions. It has also shown that stimulus generalization falls monotonically during paired-associate learning and that intrusion errors constitute a valid measure of S generalization only when initial $H_3 = 1.00$; otherwise, the IE curve tends to rise at first and then fall with continued practice.

SUMMARY

A stimulus-response analysis of paired-associate learning is described, postulating that the learning of each of the pairs involves the formation of three different connections: (*a*) between the stimulus member of the pair and a mediating, labeling response which discriminates it from the stimulus members of the other pairs; (*b*) between the stimulus produced by that labeling response and the response member of the pair; and (*c*) between the stimuli produced by each successive subelement of the response member and the following subelement. A quantitative model which relates the strength of these component habit connections to various aspects of performance on paired associates is described.

Deductions from the model were tested by means of an experiment involving 10 *S*s in each of six conditions. The antecedent variables were the difficulties of forming each of the three component habits described above, manipulated by using different stimulus members and response members in the pairs, and by assigning the responses to the stimuli in varying ways. With these materials it was possible to analyze the responses given by the *S*s so as to obtain separate measures of the strength of the three component habits during learning. A typical paired-associate learning situation was employed, using a Hull memory apparatus and the anticipatory method of performance.

It was found that the relations between the gross learning scores in the different conditions predicted from the model agreed closely with the obtained results in direction and magnitude. The obtained practice curves for intrusion errors agreed closely over a wide variety of conditions with those predicted from the model which postulates that the number of intrusions is a function of the strengths of all three of the habits described above, and not just of that of the discrimination habits as assumed in some studies. The intrusion error curves tended to first rise, then fall slightly with practice under most conditions. The practice curve for stimulus generalization during paired-associate learning was found to have a monotonically falling slope rather than the first rising, then falling shape sometimes inferred.

REFERENCES

BAKAN, D. A generalization of Sidman's results on group and individual functions and a criterion. *Psychol. Bull.*, 1954, **51**, 63–67.

BATTIG, W. F. Comment on "intralist generalization in paired-associate learning." *Psychol. Rev.*, 1959, **66**, 338–339.

BUGELSKI, B. R., & SCHARLOCK, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

GAGNÉ, R. M. The effect of sequence of presentation of similar items on the learning of paired associates. *J. exp. Psychol.*, 1950, **40**, 61–73.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GIBSON, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.

GIBSON, E. J. A re-examination of generalization. *Psychol. Rev.*, 1959, **66**, 340–342.

Gibson, J. J., & Gibson, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, 62, 32–41.

Goss, S. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, 46, 419–428.

Hake, H. W., & Ericksen, C. W. Role of response variables in recognition and identification of complex visual forms. *J. exp. Psychol.*, 1956, 52, 235–243.

Hovland, C. I. Experimental studies in rote learning theory: VII. Distribution of practice with varying lengths of lists. *J. exp. Psychol.*, 1940, 27, 271–284.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Hunt, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, 57, 262–267.

Kurtz, K. H. Discrimination of complex stimuli: The relationships of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, 50, 283–292.

Lewis, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, 12, 65–78.

McAllister, D. E. The effect of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, 46, 329–336.

Mandler, G. Response factors in human learning. *Psychol. Rev.*, 1954, 61, 235–244.

Murdock, B. B. Intra-list generalization in paired-associate learning. *Psychol. Re v.* 1958, 65, 306–314.

Murdock, B. B. A reply to Battig, Gibson, and Runquist. *Psychol. Rev.*, 1959, 66, 345–346.

Nobel, C. E., & McNeely, D. The role of meaningfulness (*m*) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, 53, 16–22.

Pishkin, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *J. exp. Psychol.*, 1960, 59, 371–378.

Runquist, W. N. Remarks on "intra-list generalization in paired-associate learning." *Psychol. Rev.*, 1959, 66, 343–344.

Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, 49, 287–293.

Sidman, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, 49, 263–269.

Underwood, B. J., & Goad, D. Studies of distributed practice: I. Influence of intra-list similarity in serial learning. *J. exp. Psychol.*, 1951, 42, 125–133.

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, 58, 70–78.

(Received July 21, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 348–356

INTERACTIONS OF SUPRATHRESHOLD TASTE STIMULI[1]

JOSEPH M. KAMEN, FRANCIS J. PILGRIM, NORMAN J. GUTMAN, AND BEVERLEY J. KROLL

Quartermaster Food and Container Institute for the Armed Forces, Chicago, Illinois

Probably everyone has some hunches on how one taste quality affects another, on how sugar affects the taste of salt, on how acid affects the perception of bitterness, and the like, but many of these assertions are contradictory. Probably some of this disagreement is due to people's confusing changes in intensity of the individual taste qualities with the more ambiguous tastes created by the addition of one unitary taste stimulus to another. Thus, it may be that a salt-and-sugar solution produces relatively unclear taste sensations even though the individual qualities, when observed with a more analytic set, have the same subjective intensities as they would have in a pure salt or pure sugar solution.

No systematic investigation of taste interactions at suprathreshold stimulus intensities has ever been reported. Anderson (1950) has presented a critical review of the literature and the results of his own systematic study on interactions among stimuli at near-threshold concentrations. Fabian and Blum (1943) have summarized the early literature on taste interactions and reported their investigation of interactions between various sweet, salty, and sour substances; however, their emphasis was on the effect of a subthreshold concentration of one substance upon the perceived intensity of a suprathreshold concentration of another. They concluded that a subthreshold concentration of salt (NaCl) increased the intensity of five different sugars and reduced the sourness of five organic acids and one inorganic, hydrochloric acid (HCl). Each subthreshold concentration of the five sugars, in turn, reduced saltiness of NaCl and the sourness of the six acids. All organic acids enhanced saltiness, although HCl appeared to have no effect; but the effect of acids upon sweetness seemed to be in part a function of the specific acids and the specific sugars used. For example, the sweetness of fructose was reduced by lactic, tartaric, acetic, and malic acids; but HCl and citric acid had no effect. In contrast, citric, lactic, and tartaric acids enhanced the sweetness of sucrose, but HCl and acetic acid seemed to have no effect.

More recently, Beebe-Center, Rogers, Atkinson, and O'Connell (1959) have reported on the interactions between suprathreshold concentrations of NaCl and sucrose. Their major conclusion was that some enhancement of sweetness by salt was evident in the case of weak solutions, but the principal effect was one of masking.

For the purposes of the present study, we assumed the existence of four basic taste qualities—salt, sweet, sour, and bitter—and that the appropriate stimulus for each is NaCl, sucrose, citric acid, and caffeine, respectively. The interactions investigated were those between every pair of qualities. In each such pair, a given stimulus was studied both as to its effect on another and how it was affected by the other. Thus, mixtures of sucrose and NaCl were examined from two points of view: the effect of sucrose (secondary stimulus) upon the perceived intensity of the saltiness of NaCl (primary stimulus); and conversely, the effect of NaCl (now the secondary stimulus) upon the perceived intensity of sweetness of sucrose (now the pri-

---

[1] This paper reports research undertaken at the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned Number 1093 in the series of papers approved for publication. The views or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the views or indorsement of the Department of Defense.

TABLE 1

PERCENTAGE CONCENTRATIONS OF SOLUTIONS USED IN STUDY OF INTERACTIONS OF TASTE QUALITIES

Taste Quality	Primary Series (Rated for Intensity)				Secondary Series (Added to Solutions)			
Salt (NaCl)	.15	.45	1.40	4.00	.00	.13	.44	1.50
Bitter (caffeine)	.031	.076	.195	.500	.00	.025	.048	.093
Sweet (sucrose)	.50	1.70	5.80	20.0	.00	.45	1.90	8.00
Sour (citric acid)	.009	.029	.089	.274	.00	.007	.023	.073

Note.—Concentrations represent weight of stimulus per 100 ml. of solution. Concentrations of citric acid have been corrected for the one molecule of water of crystallization. An exception occurred in the experiment on the effect of sucrose on saltiness. The NaCl concentrations were .10, .35, 1.20, and 4.00; and the sucrose concentrations were .00, .45, 1.70, and 6.00. This interaction experiment was the first one conducted, and the concentrations were slightly revised for the subsequent ones.

mary stimulus). In all, 12 sets of interactions are possible.

METHOD

Taste solutions.—Two series of each stimulus were prepared, a primary ("effect on") and a secondary ("effect of"). The concentrations in the primary series were intended to cover the range of intensities from barely perceptible to almost extreme. The concentrations in the secondary series were intended to result in perceived intensities from none to moderate, since it was thought that interaction effects would be most easily demonstrated if the four concentrations of the secondary series were of generally lower intensity than the primary series. Selections of the specific concentrations within these ranges were based upon previous unpublished data.

All concentrations are shown in Table 1 and represent the number of grams of the solute per 100 ml. of solution. The weights for citric acid have been corrected for the one molecule of water of crystallization per citric acid molecule. It will be noted that, apart from the 0% concentrations of the secondary stimulus, the concentrations in both series are approximately logarithmically spaced. The sucrose and citric acid were Merck Reagent, the NaCl was Merck C. P., and the caffeine was Pfizer U. S. P. Charcoal-filtered distilled water was always used as the solvent.

In each interaction experiment, the 16 solutions consisted of each concentration in a given primary series with each concentration in one of the three remaining secondary ones (Table 1). For example, in the experiment on the effect of citric acid upon the perceived intensity of bitterness, the primary stimulus series of caffeine (.031%, .076%, .195%, .500%) is used in combination with the secondary series of citric acid (.00%, .007%, .023%, .073%), as shown in Table 2.

Experimental design.—Each interaction experiment was independently replicated, the interval between replications varying from 1 wk. to 16 mo. The basic experimental design is given in Table 2, where the experiment on the effects of citric acid upon bitterness is shown as an example.

The levels of the primary stimulus, caffeine, and the secondary stimulus, citric acid, were taken from Table 1. The 16 solutions were divided into two sets. Half the judges (*O*s) in each replication evaluated the solutions marked "O" in Table 2, while the other half evaluated the solutions marked "X."[2]

Judges.—The *O*s were selected from a pool of approximately 700 civilian and military, male and female, employees who routinely participate in preference tests of foods though

---

[2] The design is a half-replicate in which the interaction of the "linear" component of the primary stimulus and the "cubic" component of the secondary is confounded with Judge-Group. Quotation marks are used to indicate that these components are not linear or cubic in the quantitative sense, but rather involve comparisons between pairs of levels. Thus, the "linear" (Component I) of caffeine, when used as the primary stimulus, means that the average perceived bitterness of the eight solutions in the first two caffeine levels is compared with the average of all eight solutions in the second two levels. Similarly, the "cubic" (Component III) of caffeine compares the average perceived intensity of the eight solutions in the first and third levels with the average of the eight solutions in the second and fourth. The quadratic (Component II) is a true quadratic and involves comparing the middle two levels against the lowest and highest ones.

TABLE 2

EXPERIMENTAL DESIGN EXAMPLE: EFFECTS OF CITRIC ACID UPON THE PERCEIVED INTENSITY OF BITTERNESS

Levels of Citric Acid (Secondary Stimulus)	Levels of Caffeine (Primary Stimulus)			
	A .031%	B .076%	C .195%	D .500%
I-.00%	X	X	O	O
II-.007%	O	O	X	X
III-.023%	X	X	O	O
IV-.073%	O	O	X	X

Note.—Solutions marked "X" were evaluated by half the *O*s, and the solutions marked "O" were evaluated by the other half. Thus, the interaction of the "linear" component of caffeine and the "cubic" component of citric acid are confounded with Judge-Group.

rarely in psychophysical investigations. Independent selections of 40 *O*s were made from the pool for each replication of each experiment. Departures from randomness occurred when some were absent or were otherwise not available on the days the tests were conducted. Because 960 persons were required (12 interactions × 40 *O*s × 2 replications) and because replacement into the pool followed selection, some *O*s participated in more than one experiment or replication.

Psychophysical method.—The single stimulus method was used with a nine-interval rating of intensity. Alternate intervals were anchored with the following descriptions of intensity: *none, slight, moderate, strong,* and *extreme.* The intervals were assigned successive integers from 1 (*none*) to 9 (*extreme*) and the ratings then treated quantitatively. The *O*s were instructed to rate the intensity of the quality represented by the primary stimulus, ignoring other qualities that might be present. They were told not to swallow the samples.

Fresh solutions were prepared for each replication session. A session was usually completed in 1 day although occasionally 2 consecutive days were needed to test the 40 *O*s. Each *O* sat in a semi-enclosed testing booth. They were presented one at a time with 6-ml. samples in coded 1-oz. glasses through a turntable in a wall separating the booth from the serving area. After rating each sample, *O* rinsed his mouth ad lib. with charcoal-filtered distilled water. The time between the rating of one solution and the presentation of the next was 30 sec. During the course of the experiments, the question arose as to whether these untrained and unscreened *O*s were aware of the characteristics of a sour or bitter substance and the distinctions between them. It was decided that on the second replication of each interaction experiment, *O* would receive a reference sample of the primary stimulus prior to rating the other eight and would be asked to note carefully its flavor without rating it. The reference sample was always a pure solution of the second highest concentration of the primary stimulus. In the analyses of variance, session was a source of variation although it is a generic term that includes ordinary session variability per se, whether or not a reference sample was served, actual differences among judge groups, etc.

RESULTS

A separate analysis of variance was performed for each taste interaction. The total variation was partitioned among the following sources of variation: each orthogonal component of the primary and of the secondary stimulus, the interaction of every component of the primary with every component of the secondary stimulus, session, interaction of session with each orthogonal component and with each primary-secondary interaction, judge, and the interaction of session and solution. Each source except the last, had 1 *df*. There were 76 *df* for Judge and 532 for Judge × Solution. The .01 level was chosen as the criterion for significance.[3]

Except for those sources of variation which were confounded with Judge-Group, the error term was Judge-Solution interaction (within groups). Those sources, for which variation among Judges (within groups) was the error term, were Session, the interaction of Component I of the secondary stimulus and Component III of the primary, and the three-factor interaction of Session × Secondary-I × Primary-III.

---

[3] Four tables showing the mean ratings of each solution and one table showing the sources of variation and their corresponding *df*'s, mean squares, and levels of significance have been deposited with the American Documentation Institute. Order Document No. 6777 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief Photoduplication service, Library of Congress.

As would be expected, Component I of the primary stimulus was in each case significant, far beyond the .001 level. In 21 of 24 cases, Components II and III were also highly significant. Inspection of the ratings[3] shows that the major effects of increasing the primary stimulus concentrations were true linear, although significant departures did occur.[4]

Each mean was based upon 40 ratings, 20 in each of the two replications. The mean ratings of each of the 16 solutions in each interaction experiment are plotted in Fig. 1. The results and conclusions will be discussed separately for each interaction. The general error term, Judge-Solution interaction, will be indicated in parentheses.

Effects of Caffeine

Upon saltiness.—No significant effects of caffeine upon saltiness were found. Regardless of the level of caffeine, saltiness was merely a function of the salt itself. There was only a slight suggestion that if the caffeine level were increased even further, saltiness might eventually be enhanced. (Error $MS = 1.39$.)

Upon sweetness.—No variables affected sweetness other than the sucrose concentrations themselves. However, Caffeine-I was almost significant; the observed F was 6.50, compared to an F of 6.64 required for significance at the .01 level. Only a suggestion of masking by caffeine was present, the difference between the lower two and higher two caffeine concentrations being only .26 scale points. Higher levels of caffeine might demonstrate eventual masking of sweetness. (Error $MS = 1.70$.)

Upon sourness.—Components I and III of caffeine were each significant at the .001 level. Inspection of the mean intensity ratings reveals that the effect was one of enhancement, with no other significant sources of variation complicating the interpretation. (Error $MS = 2.94$.)

Effects of NaCl

Upon bitterness.—No significant effects were found, other than those attributable to caffeine. The curves were somewhat jagged and suggested a Caffeine-I × Salt-III interaction. However, the interaction of these two components was confounded with Judge-Group, and hence the variation among *O*s was used as the appropriate error term. (Both Judge-Solution and between-Judge *MS*s were unusually high, 3.68 and 9.63, respectively.) Despite the fact that the same types of curves emerged in the two replications and that there was a high mean square for the interaction, use of the large error term (between-Judge rather than Judge × Solution) worked against obtaining a significant F. In view of the magnitude of the mean square for this source of variation, a logical next step would be to replicate this entire experiment with a different set of confounding relationships, e.g., Primary-III × Secondary-II. (Error $MS = 3.68$.)

Upon sweetness.—Salt Components I and II were significant at the .01 level. This is interpreted to mean that salt generally tends to mask sweetness, and there is some curvature in the effects. The interpretation is complicated by two interactions which were also significant at the .01 level: (*a*) The first is the Salt-I × Sucrose-I interaction. For the lower sucrose concentrations, the various levels of salt had relatively little effect. Instead, the reduction of sweetness

---

[4] If orthogonal polynomials are used on the mean ratings (Anderson & Bancroft, 1952), it will be found that the true linear is greater than Component I, with a corresponding decrease in the mean square for the true cubic. This effect is due to the fact that although Component I is mostly linear, it does reflect some cubic; Component III, in turn, contains some linear.

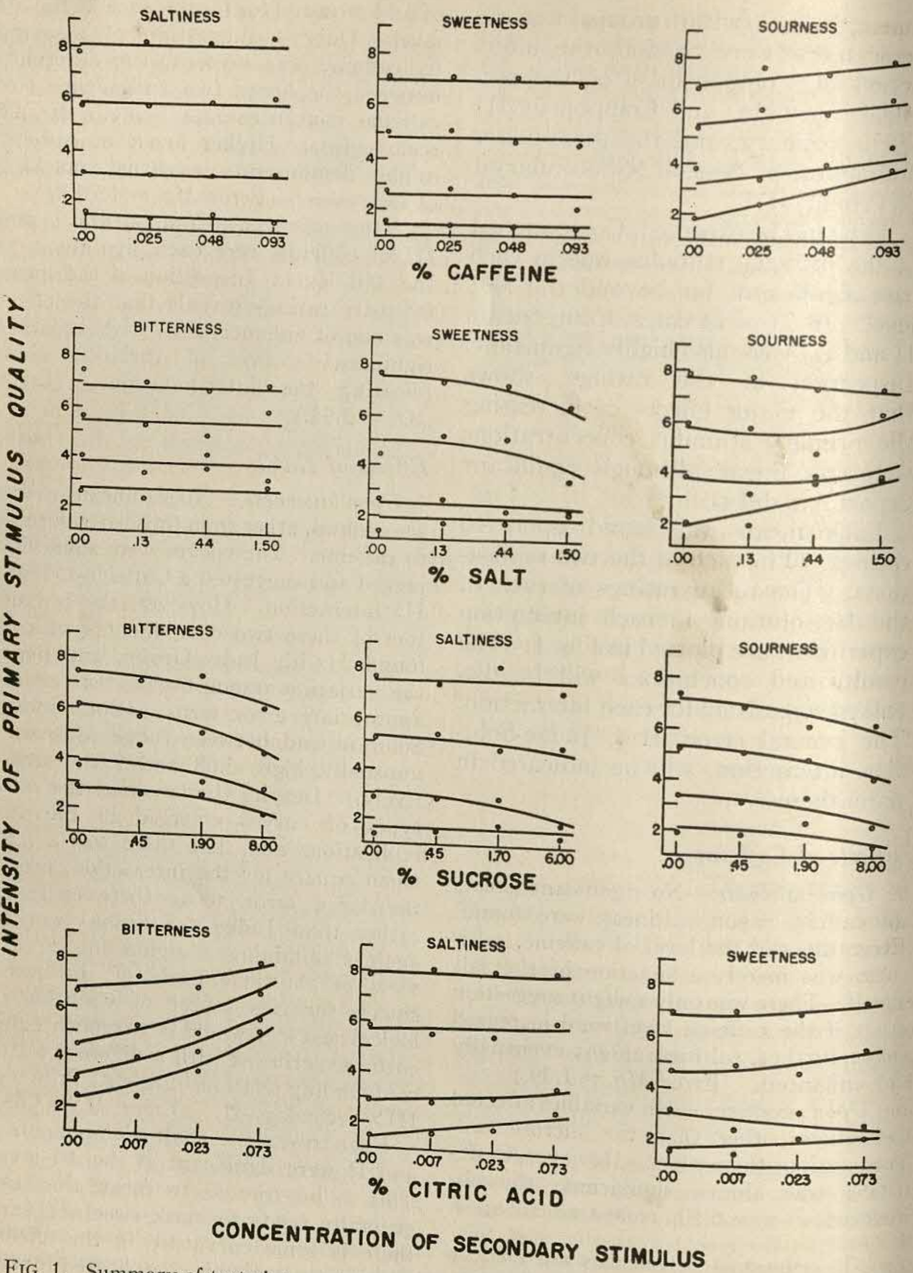


FIG. 1. Summary of taste interactions. (The abscissa represents increasing concentrations of the secondary stimulus. The four curves in each graph are for the four levels of the primary stimulus whose taste quality is shown on each graph. See Table 1 for concentrations. Curve fitting was guided by the significance of the sources of variation in the analyses of variance, rather than by least squares methods.)

occurred primarily for solutions that were high in both salt and sucrose. (*b*) The second is the Salt-II × Sucrose-I interaction. Again, for the lower sucrose concentrations, the various levels of salt had relatively little differential effect on sweetness. The greatest effect was on the higher sucrose concentrations by the highest salt concentrations, with a lesser effect by the lower salt concentrations.

In both of these interactions, the very highest salt concentration had its greatest sweetness-depressing effects at the higher sucrose concentrations. Thus, while the over-all effect of salt upon sweetness was one of masking, this did not occur or was not as pronounced for the lowest sucrose concentrations. In fact, for the very lowest sucrose, salt seemed to enhance sweetness.

The results are consistent with the conclusion of Beebe-Center et al. (1959), that some enhancement of sweetness by salt does occur although the major effect was one of masking, and with the report by Fabian and Blum (1943) that near-threshold concentrations of salt enhance sweetness.

Future research might well be devoted to more intensive study, at just above threshold salt concentrations, of sucrose concentrations between .5% and 6%, the region in which the shift from enhancing to masking appears to occur. (Error $MS = 2.24$.)

Upon sourness.—The results of this experiment are more complex than those of the others. Apart from the citric acid itself, Component-II only of salt had an effect ($P < .01$); the highest and lowest (0%) salt concentrations had less effect than the middle ones. This result, however, should be viewed in the light of the significance of two interaction terms: Salt-I × Citric Acid-I ($P < .001$); Salt-I × Citric Acid-III ($P < .01$). The latter may have arisen because Component III has some linear effects. The highest levels of salt tended to enhance the sourness of the lower concentrations of citric acid, but reduced the sourness of the higher acid concentrations.

From these two significant sources of variation, from the fact that they appeared in both sessions, from the absence of other significant effects (except between-judge variation), and from inspection of the mean ratings, we can infer that low level salt depressed sourness, but high levels enhanced it so that an over-all effect was not apparent. The higher the acid concentrations the later these two stages appear—in terms of increasing salt concentrations. Thus, .13% salt appeared to have reduced the sourness of the lowest two acid concentrations, but as the salt was increased to .44%, enhancement took place; .44% salt had a depressing effect on the highest two levels of acid, but 1.50% salt seemed to increase sourness.

No explanation of this nonmonotonic function is apparent. Quantitative chemical analyses of the solutions failed to reveal errors in making them up, nor did examination of the ratios of the normality of the salt to the normality of the acid suggest a chemical explanation. What is needed is an extended range of salt, a more detailed study of the points around the minimum sourness intensities produced by salt at each citric acid concentration, and use of other acids as sour stimuli. (Error $MS = 2.74$.)

Effects of Sucrose

Upon bitterness.—Sucrose reduced the intensity of bitterness. Component I was significant at the .001 level, while Components II and III were significant at the .01 level. Although successively increasing sucrose concentrations consistently reduced bitterness, the highest concentration seemed to have a disproportionately large effect.

Two solutions (.45% sucrose, .076% caffeine; 1.9% sucrose, .50% caffeine) appeared to deviate somewhat from the downward trend. The significance ($P < .001$) of the Primary-III × Secondary-III × Session interaction largely reflects this deviancy and the fact that it occurred for only one session. Hence, its bearing on the major conclusion is negligible.

The Caffeine-I × Session interaction was also significant ($P < .001$), and was due to the fact that *O*s in Session 1 did not use the upper categories of the rating scale as often as did *O*s in Session 2; thus, the range of their average ratings was lower. This phenomenon, commonly appearing during psychophysical tests, is not considered important.

Except for the two deviant solutions on one session, the effects of sucrose were fairly clear-cut—sucrose reduced bitterness. (Error $MS = 2.74$.)

Upon saltiness.—Sucrose had no general enhancing or masking effects on saltiness, a result in agreement with that reported by Beebe-Center et al. (1959). Also in agreement was the indication that the relationship might be somewhat complex. Thus, the interaction of Salt-III with Sucrose-III was significant ($P<.01$); and inspection of the mean ratings suggests that this may be attributable to the highest sucrose concentration (and to a lesser extent, the next-to-lowest sucrose concentration) rather sharply reducing the saltiness of the highest and next-to-the-lowest salt concentrations. Why this effect did not occur for the second highest salt concentration is not readily apparent.

The Salt-II × Sucrose-I interaction was also significant ($P < .01$), but this effect seemed to occur primarily for one session. Because the triple interaction involving Session was also significant ($P < .001$), not much importance is attributed to the significance of the simple interaction.

Apart from the absence of over-all enhancing or masking effects, the results of this experiment are not as definitive as those of the others. Further research should be devoted not only in replicating this one, but should also extend the sucrose concentrations to perhaps 15 or 20%. (Error $MS = 1.69$.)

Upon sourness.—All three components of sucrose were significant, Component II at the .01 level and the others at the .001 level. Sucrose reduced the intensity of sourness, a result in general agreement with that reported by Fabian and Blum (1943) and those cited by them. The sharpest drop in intensity occurred for the 6.00% sucrose concentration. If the range of sucrose concentrations were extended, the masking effects would probably be even more pronounced. (Error $MS = 2.60$.)

Effects of Citric Acid

Upon bitterness.—Citric acid very markedly enhanced bitterness, as is demonstrated by the significance of all three components, I and III at the .001 level, Component II at the .01 level. The steepest rise in bitterness was evident between the highest and next highest citric acid concentrations. The Citric Acid-I × Caffeine-I interaction was also significant ($P < .001$). The acid had a proportionately greater effect upon the lower concentrations of caffeine than on the higher concentrations. This effect may be partly at least due to restriction of the rating scale at the highest caffeine concentrations.

The Caffeine-I × Session interaction was also significant ($P < .001$). In Session 2, *O*s used a narrower range in evaluating the solutions. No special importance is attributed to this result.

The conclusions from this experiment are probably so clear that further study would not be as fruitful as with several of the other interactions. (Error $MS = 2.79$.)

Upon saltiness.—Saltiness was generally enhanced by citric acid, as shown by the significance ($P < .001$) of the Citric Acid-I component. The enhancement has not been shown to be dependent upon the level of salt. The absolute increase was not very marked, being only .32 scale points between the lowest two and highest two citric acid concentrations; but, the error term is the lowest of all interaction experiments.

Salt-I × Session interaction was also significant at the .001 level. The *O*s in Session 2 tended to restrict their range of ratings, this restriction being manifested more at the higher levels. As in the previous interaction, this finding is not of any special importance.

Extending the range of citric acid

concentrations might reveal a drop-off in enhancement, particularly for the lower salt concentrations, and perhaps an eventual masking for the higher salt concentrations. (Error $MS = 1.47$.)

Upon sweetness.—Citric acid generally increased sweetness ($P < .01$). This major conclusion is in agreement with that of Fabian and Blum (1943), who used only near-threshold concentrations of citric acid. The only other significant ($P < .01$) source of variation, Primary-I × Session, means that in Session 1, *O*s rated the solutions containing the weakest concentrations of sucrose higher than did *O*s in Session 2. Thus, the range of ratings was more restricted on Session 1 than on Session 2.

Extending the range of citric acid concentrations should aid in better defining the mathematical relationship between citric acid and perceived sweetness, particularly in determining whether the increase finally levels off and perhaps changes to a decrease. (Error $MS = 1.98$.)

DISCUSSION

In most experiments the results were clear-cut, and the functional relationships between the primary and secondary stimuli were either monotonic or nonexistent. Where ambiguities or hints of a trend with different stimulus concentrations appeared, recommendations for follow-up research were indicated. No secondary stimulus had a uniformly enhancing or depressing effect on the remaining three primaries; nor was any primary uniformly enhanced or depressed by other secondaries. Also, what happened at near-threshold stimulus concentrations was not necessarily predictive of suprathreshold phenomena.

The general results from the "linear" comparisons may be summarized as follows: (*a*) Caffeine does not appear to affect saltiness, nor does salt appear to affect bitterness. (*b*) Caffeine does not seem to increase or decrease sweetness, but sucrose depresses the perceived intensity of bitterness. (*c*) Caffeine and citric acid have a mutually enhancing effect upon the taste quality specific to each. (*d*) Salt decreases sweetness, but sucrose does not appear to affect saltiness. (*e*) Salt seems to have no monotonic effect upon sourness, but citric acid increases saltiness. (*f*) Sucrose decreases sourness, but citric acid enhances sweetness.

Conspicuously absent from this paper is reference to physiological correlates of taste perception. The assumption that four primary taste qualities exist does not imply that four types of receptors also exist. Pfaffmann (1958), in his study of electrical recording of nerve impulses from single taste nerve fibers of the rat, was unable to find complete specificity of receptor action. For example, NaCl and sugar activated the same sensory nerve fiber and its attached sense endings. The possible afferent discharge patterns, Pfaffmann concluded, may include not only an increase but a decrease in neural flow and that the primary taste qualities represent nodal points in the manifold of taste sensations rather than basic receptor types.

The taste interactions reported here probably have a neurological basis, and the complexity of some of the results of the interaction experiments (e.g., NaCl-sucrose interactions) may reflect complex neural patterns such as those described by Pfaffmann.

The design, method, and *O*s used here were not typical of those used in most psychophysical research. One intrinsic disadvantage of the half-replicate design with specified confounding relationships is that no *O* evaluates all solutions, so that the error terms for a few sources of variation are larger than in a full factorial design and so that irregularities in curves can be attributed either to differences in the two groups or to "real" effects. Balanced against this weak point is the efficiency of the design and the single stimulus method (cf., matching method) in minimizing testing time of *O*s and in avoiding loss of motivation through testing of all solutions at one sitting. It was not feasible to have the same *O*s return to a second session to complete the ratings of the eight samples they did not test during the first.

Nevertheless, in view of suggestions of possible significance of the variable which was confounded with Judge-Group, a different confounding relationship might have proved to be more revealing.

The relative difficulty in rating—the ambiguity of the sensations—is partially reflected in the magnitude of the error terms. Saltiness seemed to be the easiest to evaluate, insofar as *O*s tended to agree more with each other on this quality than on the others. Sweetness was next lowest, but bitterness and sourness had errors of the order of twice that of saltiness. A design similar to the one used here would be too laborious to employ for interactions among three or four stimuli, but the present data should prove useful in selecting the optimum concentrations for exploring the relationships among mixtures of more than two stimuli.

SUMMARY

Twelve experiments were conducted to determine how each unitary taste quality is affected by each of the other taste qualities. Solutions containing stimuli appropriate to both taste qualities were rated for intensity of sweetness, saltiness, sourness, or bitterness by a group of judges. Each experiment was independently replicated. In most cases, the effects were those of simple enhancement or masking, or no effect at all was found. Certain exceptions and complex relationships occurred, and recommendations for follow-up research were made. Various aspects of the method and design were discussed.

REFERENCES

ANDERSON, R. J. Taste thresholds in stimulus mixtures. Unpublished doctoral dissertation, University of Michigan, 1950.

ANDERSON, R. L., & BANCROFT, J. A. *Statistical theory in research.* New York: McGraw-Hill, 1952.

BEEBE-CENTER, J. G., ROGERS, M. S., ATKINSON, W. H., & O'CONNELL, D. N. Sweetness and saltiness of compound solutions of sucrose and NaCl as a function of concentration of solutes. *J. exp. Psychol.*, 1959, **57**, 231–234.

FABIAN, F. W., & BLUM, H. B. Relative taste potency of some basic food constituents and their competitive and compensatory action. *Food Res.*, 1943, **8**, 179–193.

PFAFFMANN, C. Behavioral responses to taste and odor stimuli. In, *Flavor research and food acceptance.* New York: Reinhold, 1958.

(Received July 22, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 357–360

DIFFERENTIAL COST, GAIN, AND RELATIVE FREQUENCY OF REWARD IN A SEQUENTIAL CHOICE SITUATION[1]

JEROME L. MYERS, RAYMOND E. REILLY, AND HARVEY A. TAUB

University of Massachusetts

In a recent experiment Taub and Myers (1961) studied the effects of differential gain in a two-choice situation. The gain received by *S* depended on which of two events was correctly predicted (the less frequent event paid more); the cost of an incorrect prediction was the same, regardless of which event had been predicted. The *S*s did not approximate the optimal strategy[2] of always predicting the event associated with the higher expected value (*EV*). That *S*s could discriminate between *EV*s was indicated by two results: (*a*) the event with the higher *EV* was predicted more than 50% of the time in all combinations of conditions, and (*b*) the greater the difference in *EV*s (ΔEV) for two events, the greater was the difference in percentage prediction of the two events.

The present study extends Taub and Myers' investigation by introducing the variable of cost ratio, i.e., the ratio of the loss associated with an incorrect prediction of one event relative to the loss associated with an incorrect prediction of the alternative event. Interactions among the effects of cost, gain, and frequency are of special interest, as are the relative effects of gain and cost upon performance. The relation of choice behavior to the *EV*s associated with events will be considered.

METHOD

Apparatus.—The apparatus has been described in detail elsewhere (Taub & Myers, 1961). Four panels, each 8½ in. high and 7 in. wide were mounted at a 45° angle away from *S*. Each panel contained four lights and two switches, arranged in two columns. The upper two lights were green and constituted the informing lights. The lower two lights were amber and indicated *S*'s choice on each trial. The choice lights were controlled by the two switches operated by *S*.

The informing light sequence was constructed from a table of random numbers subject to the following restrictions: (*a*) each light appeared a designated percentage of the time; (*b*) only one light was correct on each trial. The sequence was programed by a Western Union Tape Transmitter.

Subjects.—The *S*s were 216 undergraduate volunteers attending the University of Massachusetts. None had previously participated in an experiment of this sort.

Procedure.—Each *S* was given 100 chips, worth ½¢ each, at the beginning of the session. The *S*s were instructed in the use of the choice switches, and in the purpose of the experiment. They were told how much a correct choice of either light would pay, and how much an incorrect choice of either light would cost. Chips were cashed in at the end of 150 trials.

The *S*s were assigned randomly to 27 groups of 8, each serving under a different combination of relative frequency, gain, and cost ratios. The relative frequencies (percentages of occurrence of the two informing lights) were 90–10, 70–30, and 50–50. Gain ratios were 1 chip to 1 (1:1), 1 chip to 2 (1:2), and 1 chip to 4 (1:4). Cost ratios were the same as these. In all cases the light with the lower frequency of occurrence yielded the higher gain and the lower cost. Under the 50–50 conditions the light with the higher gain had the lower cost. In each of the 27

---

[1] This experiment was supported by funds provided by the United States Navy Training Device Center, Port Washington, New York, under Contract 61339-588.

[2] Optimal strategy here refers to that sequence of choices which maximizes the expected monetary payoff.

cells, 4 *S*s saw the low gain light in the right panel position, and 4 *S*s saw it in the left panel position.

RESULTS AND DISCUSSION

The mean percentages of prediction in the last 50 trials of the high frequency, high cost, 1-chip gain light (E_1) are presented in the column labeled *p* in Table 1. The number of choices of an event increases as relative frequency increases, as cost decreases, and as the gain associated with the alternative decreases. The data from the 70–30 and 90–10 groups involving 1:1 gain and cost ratios suggest that Estes' matching solution for the noncontingent case is unsatisfactory under monetary incentive conditions. Although this solution has been verified in a number of experiments not involving monetary incentives (e.g., Estes & Straughan, 1954; Grant, Hake, & Hornseth, 1951), several other studies using incentive have resulted in data similar to those of the two groups under discussion in the present study (Edwards, 1956; Goodnow, 1955; Siegel & Goldstein, 1959; Taub & Myers, 1961). All these studies which used incentive obtained *p* values which were close together at comparable levels of relative frequencies. For example, Goodnow, Edwards, and the present investigators all report *p* values for $\pi = .70$, (for 100 to 150 trials) between .80 and .83. Thus it appears that even the small monetary incentives used in the present study are sufficient to evoke choice probabilities above matching. An alternative explanation is that the experiments cited provided not only incentive, but also a form of feedback (increasing and decreasing piles of chips or coins) lacking in experiments which yielded matching. The relative contributions of these two factors of motivation and feedback to choice behavior remain to be explored.

TABLE 1

EXPECTED VALUE (*EV*), DIFFERENCES IN EXPECTED VALUE (Δ*EV*), AND PERCENTAGE CHOICE OF E_1 IN THE LAST 50 TRIALS (*p*)

Cost Ratio	Gain Ratio	Relative Frequency	*EV*	Δ*EV*	*p*
1:1	1:1	50–50	.00	.00	55.25
1:1	1:1	70–30	.40	.80	80.25
1:1	1:1	90–10	.80	1.60	96.25
1:1	1:2	50–50	.00	−.50	35.25
1:1	1:2	70–30	.40	.50	63.50
1:1	1:2	90–10	.80	1.50	95.75
1:1	1:4	50–50	.00	−1.50	32.75
1:1	1:4	70–30	.40	−.10	45.75
1:1	1:4	90–10	.80	1.30	86.00
1:2	1:1	50–50	−.50	−.50	27.75
1:2	1:1	70–30	.10	.50	76.25
1:2	1:1	90–10	.70	1.50	91.25
1:2	1:2	50–50	−.50	−1.00	22.50
1:2	1:2	70–30	.10	.20	59.75
1:2	1:2	90–10	.70	1.40	90.25
1:2	1:4	50–50	−.50	−2.00	33.75
1:2	1:4	70–30	.10	−.40	54.25
1:2	1:4	90–10	.70	1.20	91.50
1:4	1:1	50–50	−1.50	−1.50	11.00
1:4	1:1	70–30	−.56	−.10	65.25
1:4	1:1	90–10	.50	1.30	91.00
1:4	1:2	50–50	−1.50	−2.00	17.75
1:4	1:2	70–30	−.50	−.40	46.75
1:4	1:2	90–10	.50	1.20	89.00
1:4	1:4	50–50	−1.50	−3.00	15.50
1:4	1:4	70–30	−.50	−1.00	32.50
1:4	1:4	90–10	.50	1.00	81.75

The optimal solutions, in terms of maximizing monetary payoffs, would be always to predict the light with the higher *EV*. At the end of 150 trials, few *S*s are utilizing this solution. The 50–50 groups' failure to approach this solution more closely is particularly difficult to account for. Each of eight groups is faced with two equally frequent events, one of which pays as much or more and loses the same or less than its alternative. Yet

TABLE 2

ANALYSIS OF VARIANCE OF THE FREQUENCY OF CHOICE OF E_1

Source	*df*	*MS*	*F*
Relative frequency (F)	2	17,489.24	440.76***
Cost (C)	2	1,298.44	32.72***
Gain (G)	2	823.33	20.75***
F × C	4	119.48	3.01**
F × G	4	323.04	8.14***
C × G	4	54.56	1.38
F × C × G	8	80.65	2.03*
Error	189	39.68	—

* $P < .05$.
** $P < .025$.
*** $P < .001$.

percentage choice of the low-*EV* light ranges from 10.5 to 35.5. It is not clear whether *S*s expect the higher cost light to occur more frequently or whether they perceive the situation correctly but are incapable of reaching a rational decision, at least in 150 trials.

The results of an analysis of variance of the frequency of choice appear in Table 2. All results were significant at the .01 level except Cost × Gain ($P > .05$), Cost × Frequency ($P < .025$), and Cost × Gain × Frequency ($P < .05$).

The extent and direction of main and interaction effects can be predicted by assuming a linear relation between ΔEV and p. If the mean ΔEV for the three levels of relative frequency of gain, and of cost are computed from the ΔEV column of Table 1, it is evident that (*a*) differences should exist among the mean p values for the levels of each of these variables, (*b*) gain and cost should contribute about equally to the total variance (i.e., the variances of the mean ΔEV for these two variables are the same), and (*c*) frequency should contribute about nine times as much variability to the data matrix as should gain or cost. Actually, cost seems to have a somewhat greater effect than gain, and the frequency mean square is larger in relation to those of cost and gain than the variance of the ΔEVs would suggest. The mean ΔEVs for cost-frequency combinations yield the prediction that as the cost associated with E_1 increases, p should decrease, and that this decrease should be greater as relative frequency decreases. This in fact occurs, the greatest spread among the mean p values for level of cost being observed for the 50–50 relative frequency condition. A similar prediction can be made for the Gain × Frequency interaction, that as the gain associated with the alternative increases, p should decrease, and that this decrease should be greater as relative frequency decreases. The 70–30 frequency condition does show a greater variability among the gain means than does the 90–10 condition, but the three gain means are almost identical for the 50–50 condition. Comparing the results for the three cost ratios with those for the three gain ratios at 50–50 relative frequency leads to the conclusion that, at this level of frequency, increased cost associated with E_1 is more effective in decreasing choice of E_1 than is increased gain associated with the alternative. Table 1 substantiates this inference. Note these pairs of conditions: the 1:1 cost, 1:2 gain, 50–50 condition, and the 1:2 cost, 1:1 gain, 50–50 condition; the 1:1 cost, 1:4 gain, 50–50 condition, and the 1:4 cost, 1:1 gain, 50–50 condition; the 1:2 cost, 1:4 gain, 50–50 condition, and the 1:4 cost, 1:2 gain, 50–50 condition. Although the *EV*s are the same for the members of each pair, increased cost associated with E_1 yields a lower p value than increased gain on the alternative, in all three instances.

Differences in *EV* do partially account for the rank order of p values and for some main and interaction effects. However, the authors are not proposing a theory of choice behavior based on the assumption of a relationship between ΔEV and p. While the simple and familiar concept of *EV* does account for many relationships among variables and, in the absence of any more successful formal model, provides some frame of reference within which to organize the data, it is evident tht ΔEV does not sufficiently account for the data obtained. The results of various 50–50 conditions are particularly deviant since the data consistently suggest that when two events occur equally often, the tendency to choose the one resulting in the smaller loss is much greater than the tendency to choose the event resulting in the larger gain.

SUMMARY

Gain, cost, and frequency ratios were manipulated for two events in a sequential choice situation with monetary incentives. Percentage prediction of the events was strongly influenced by these variables and by their interaction. The relation between percentage prediction of an event and the expected value associated with it was discussed.

REFERENCES

EDWARDS, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GOODNOW, J. J. Determinants of choice distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

GRANT, D. A., HAKE, H., & HORNSETH, W. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychcl.*, 1959, **57**, 37–42.

TAUB, H. A., & MYERS, J. L. Differential monetary gains in a two-choice situation. *J. exp. Psychol.*, 1961, **61**, 157–162.

(Received July 23, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 361–367

THE EFFECT OF RECALL UPON RECOGNITION [1]

NELSON G. HANAWALT AND ARLENE G. TARR [2]

Rutgers University

Belbin (1950) published a paper which showed rather dramatically that recall had a depressing effect upon recognition. These experiments were inspired by a paper by Postman, Jenkins, and Postman (1948) in which recall and recognition were compared.

Postman et al. (1948) gave two groups of *S*s six trials on a list of 48 nonsense syllables. The syllables were read in random order for the different trials with instructions to remember as many as possible. Immediately after learning Group I took a recognition test followed by a 10-min. recall period. Group II had the tests in the reverse order. Recognition first had a facilitating effect upon recall, which did not surprise the authors, but they were not prepared for the fact that recall first depressed the score on the recognition test. The above experiment was not designed to test the effect of recall upon recognition, consequently the lapsed time for the two recognition tests was not controlled. The authors assumed that the lower score for the recognition test after 10 min. of recall was due to weak traces becoming weaker during the 10-min. delay.

Belbin's (1950) experiments were designed specifically to test the effect of recall upon recognition. The learning was incidental and the method that of interpolated recall. In the first experiment 64 *S*s spent 2 min. in a waiting room with a one-incident safety poster facing them on an otherwise blank wall, then half of the *S*s took a "prompted by standardized questions" recall test, followed immediately by a recognition test. The control group engaged in an unrelated activity following the viewing, and took the recognition test after the same lapse of time. The recognition test was simply a presentation of a duplicate of the original poster with instructions to say yes or no as to whether or not it was the same picture, regardless of the caption. For both Groups I and II, half of them saw a picture in recognition with a different caption. The caption made a difference, but under both conditions *S*s who had recalled, overwhelmingly said that it was not the same, whereas the control group for the most part identified it correctly. In another experiment Belbin showed that the effect was still present after 3 hr. It was concluded that the chief factors responsible for the results were errors in recall—importations and omissions. On the recognition test, the absence of some erroneously recalled detail or the presence of some nonrecalled detail seemed to determine the experimental group's rejection of the picture as being the same.

Kay and Skemp (1956), in some preliminary studies, repeated Belbin's experiment with five groups of *S*s, confirming her findings but with much less dramatic results. Experiment I was similar to Belbin's excepting that a complex picture (a many-incident park scene) was used in the learning period. Belbin's type of recognition test of the picture as a whole was used but a 30-item questionnaire based upon the incidents in the picture was added. Just how closely Belbin's design was followed is not too clear but apparently a free written recall period was substituted for the "standardized questions." The recall period came 10 min. after the 40-sec. viewing of the picture, with the recognition test coming 40 min. later. With this material and method, the Belbin effect practically disappeared for the identification of the picture as a whole, since all but 2 of the *S*s who had recalled said it was the same. However, confirmatory evidence for the depressing effect of recall upon recognition was found in the 30-item questionnaire. On the 10 items frequently recalled, the recall group did a little better, but on the 20 items less frequently recalled, the nonrecall group did nearly twice as well.

In Exp. II Kay and Skemp (1956) compared recalled vs. nonrecalled items of the same group of *S*s rather than the performance of a recall vs. a nonrecall group. In a study with outline faces, preliminary to Exp. II, they showed that the number of reconstruction errors was an important factor in recognition. However, as a result of Exp. II, which will not be presented here because the writers are unable to understand the exact procedure used, a new hypothesis was constructed, namely, "that the juxtaposition of better and worse known items will raise the threshold of recognition of the latter" (p. 159). In Exp. III some evidence in support of this hypothesis was reported but the authors admit that the experiment was not well controlled.

---

[1] This research was supported in part by the Research Council of Rutgers University.

[2] The junior author is now at the University of Chicago. The authors wish to express their appreciation to the following for assistance: Helen J. Ayars and Rose Marie Rayder for assistance in the replication of the main experiment, and to Barbara A. Kinney and Barbara S. Wich for assistance in the replication of the Postman, Jenkins, and Postman (1948) experiment.

The present experiment was designed to test the effect of interpolated recall upon recognition with delayed recognition, different material, and an improved recognition test. It seemed well established that a depressing effect was present when recognition followed immediately after recall but there was a question about the effect upon delayed recognition.

METHOD

The original experiment was conducted in 1958 with four groups of *S*s, two experimental and two control, designated below as Sample A. In 1960 a replication was run with four additional groups designated as Sample B. All *S*s took a true-false test as the learning task. Immediately following learning, Groups IA and IB and IIIA and IIIB, the experimental groups, were allowed 8 min. to recall as many as possible of the final adjectives on the T-F test. During the same time Groups IIA and IIB, and IVA and IVB, the control groups, participated in the usual class activity (lecture). Both samples of Groups I and II took the recognition test 8 min. after learning; Groups IIIA and IVA took the recognition test 48 hr. after learning; and, because of a change of schedule of classes between 1958 and 1960, the B groups could not be tested after 48 hr., consequently Groups IIIB and IVB took their recognition test 52 hr. after learning.

Subjects.—There were 102 *S*s in the original and 83 in the replication experiment, all female and taking general psychology at Douglass College. The sections were assigned to different places in the design on the basis of convenience in testing since it was desirable to have the recognition tests come on the same day to reduce the possibility of discussion of the design of the experiment. Neither of the 2 *E*s had anything to do with the teaching of the classes.

Learning task.—A learning device was desired which would appear to be a completed task since two of the groups would not be doing anything with the material immediately after the learning period—Groups IIA and IIB for 8 min., Group IVA for 48 hr., and Group IVB for 52 hr. A life-like situation was also desired. The test consisted of 23 statements with a subject, copula, and final predicate adjective, e.g., "Hotel steaks are big," or "Brown eggs are expensive." The statements were all ambiguous in order to minimize the effect of true or false, and to emphasize the final adjective which was to be recalled. The recall instructions were to write down as many as possible of the final adjectives of the 23 statements on the back of the test paper. As far as *S*s were concerned, the T-F test was the end of the experiment.

Recognition test.—The recognition test contained only a five-choice test for the final adjective, numbered in the same order as the T-F test. A synonym of the correct choice and an opposite composed two of the choices. The instructions follow:

> We will now have a recognition test. From each group of five choices select the word which appeared at the end of each sentence and place its letter in the space provided. The choice groups are in the same order as they appeared in the sentences. If you are not sure which word is correct, guess. Answer every question.

RESULTS

A comparison of the recognition scores of the experimental and the control groups is presented in Table 1. There is no evidence of a depressing effect of recall upon recognition since the recall groups all produced higher mean recognition scores than their respective control groups. After 8 min. of recall Group IA had a significantly higher recognition score than IIA, but this difference did not reach significance in Sample B. The combined Samples A and B maintained a significant difference at the 5% level. When recognition was delayed 48 or 52 hr. there was a facilitation effect of recall upon recognition as shown by the significantly higher recognition scores of both

TABLE 1

MEAN RECOGNITION SCORES

Experimental (Recall)					Control (Nonrecall)					t
Group	Test Interval	N	Mean	SD	Group	Test Interval	N	Mean	SD	
IA	8 min.	29	15.86	2.59	IIA	8 min.	26	14.04	2.44	2.64**
IB	8 min.	22	16.05	2.35	IIB	8 min.	22	15.77	2.48	.38
IA+B		51	15.94	2.50	A+B		48	14.83	2.60	2.18*
IIIA	48 hr.	24	13.54	3.26	IVA	48 hr.	23	9.54	2.54	4.97***
IIIB	52 hr.	16	11.75	3.05	IVB	52 hr.	23	7.65	2.79	4.20***
IIIA+B		40	12.87	3.26	A+B		46	8.24	2.95	6.91***

* $P < .05$.
** $P < .02$.
*** $P < .001$.

experimental samples over their respective control samples. After 52 hr. the mean recognition scores were lower than after 48 hr. but the size of the difference between experimental and control groups was little changed. It is unclear from the present data whether or not the facilitating effect changes in degree between 48 and 52 hr.

Upon analysis of the results of their Exp. I, Kay and Skemp (1956) found that the superiority of their control group on the recognition test was limited to the 20 items not frequently recalled by their experimental group. A similar analysis was made of the present data. The 11 most frequently recalled words (65% recall) were compared with the 11 least frequently recalled (24% recall). In order to make this comparison the recall frequency was determined for each of the 23 words. The *S*s were then given two scores on recognition—one score for the words with a recall value above the median, and another score on the words with a recall value below the median. The median word was not scored. The percentage of recall of all the words was 44% for Sample A and 46% for Sample B, considerably greater than the 24% of Kay and Skemp's experiment, but favorable for testing their hypothesis concerning the juxtaposition of better and worse known items. According to this hypothesis the recall group would be expected to do poorly on the less frequently recalled words.

The percentage of recognition, Sample A and B combined, of the two 11-word lists, and the total list is presented in Table 2. It is interesting

TABLE 2

PERCENTAGE OF RECOGNITION ON THE TEST AS A WHOLE COMPARED TO THE 11 MOST FREQUENTLY AND THE 11 LEAST FREQUENTLY RECALLED WORDS

Word Groups	After 8 min.		After 48 or 52 hr.	
	I (Recall) $N=51$	II (Non-recall) $N=48$	III (Recall) $N=40$	IV (Non-recall) $N=46$
All Words	69%	64%	56%	36%
Most frequently recalled	74%	64%	65%	38%
Least frequently recalled	62%	61%	44%	36%

Note.—The percentages are proportions of scores rather than of discrete individuals, consequently the significance of the differences was based upon mean recognition scores and *SD*s. Percentages are used in the table in order to compare the results with those of Kay and Skemp (1956).

to note that the percentage of recognition for Group II on all words was the same as for Kay and Skemp's (1956) comparable group (64%), but that the results were reversed for Group I, 69% to their 43.5%. On the frequently recalled words the relative difference in recognition scores was similar in their experiment and in the present one, with the recall group doing better in both cases (88.2% to 80.6% compared to 74% and 64% of the present experiment). Groups I and II showed no significant difference on the infrequently recalled words, whereas Kay and Skemp (1956) found that their nonrecall group did much better over these items (50.6% compared to the recall group's 29.1%). It should be noted, also in disagreement with their results, that the control group of the present experiment did no better on the frequently recalled words than on the infrequently recalled words at the 8-min. interval.

With delay in recognition (right half of Table 2), the recall group was significantly better than the nonrecall group on all the words as well as for the frequently and infrequently recalled words. There is no evidence in Table 2 to indicate that recall has a depressing effect upon recognition, or that the better known items inhibit the recognition of worse known items.

There is evidence for the facilitating effect of recall upon recognition which is centered on the frequently recalled words when recognition came 8 min. after recall, but a facilitating effect is present also on the infrequently recalled words after 48 to 52 hr. The 24% correct recall apparently allowed for the strengthening of enough traces to provide a margin of superiority for this group on the recognition test over the nonrecall group which had only the traces of the incidental learning as a basis for response. Since a chance recognition score is 20%, the nonrecall group apparently had few effective traces after the lapse of this longer period of time.

The pattern of correlation coefficients between recall and recognition adds further evidence to the increased influence of recall upon recognition with the lapse of time. When recognition followed immediately after recall the rho was .49 for Group IA and .46 for Group IB. When recognition was delayed for 48 hr., Group IIIA, the rho was .60, and for Group IIIB, after 52 hr., it was .70.

Belbin (1950) and Kay and Skemp (1956) have demonstrated the effect of false recall upon recognition, consequently a detailed analysis was made of the present data in this respect, using Groups I and III and combining Samples A and B. In all there were 201 false recalls classified as follows: 66% synonyms, 19% intrusions from the beginning of the T-F sentences, 9% opposites, and 6% unclassified. Fortunately 123 (61%) of the false recalls were on the recognition test as possible choices. This was due largely to the fact that a common synonym and an opposite of the correct word had been included in each multiple choice group. Thus for the 123 words the *S*s had two learned responses—the correct one learned in the T-F test and a false one learned in recall. Which would they choose in recognition, the false recall or the correct word?

The results of the above analysis are presented in Table 3. In recognition immediately following the recall period, the false recall had two chances out of three of being selected, and 48 or 52 hr. later three out of four chances. Since 80% of the total recall was correct, and only 61% of

TABLE 3

RECOGNITION RESPONSES OF THE FALSE RECALLS WHICH WERE ALSO A CHOICE ON THE RECOGNITION TEST

Group	Recogition Interval	Chose False Recall	Chose Correct Response	Chose Another Response	Number of Responses
I(A + B)	8 min.	67%	33%	0%	66
III(A + B)	48–52 hr.	74%	18%	8%	57

the incorrect recall was directly effective because of the nature of the recognition test, the total effect of recall was positive for the present experiment. If erroneous recalls were more frequent than correct recalls, and if they had a chance to function directly on the recognition test, a depressing effect of recall upon recognition would be a reasonable outcome.

The experiment of Postman et al. (1948) replicated with control for time interval.—Since Belbin (1950) and Kay and Skemp (1956) interpreted the results of the experiment of Postman et al. as evidence for the depressing effect of recall upon recognition, it was decided to repeat it with the time interval between learning and recognition equalized. Otherwise the experiment was replicated as closely as possible. The experimental (recall) and the control (nonrecall) groups were selected at random from an advanced class in psychology composed of junior and senior college women. There were 16 *S*s in each group compared to 35 for Postman et al. (1948). Following the learing period (described above), the experimental group had 10 min. for recall of the nonsense syllables; during a similar period the control group worked on pencil and paper mazes. Both groups took the recognition test 10 min. after the last learning trial.

The experimental group had a mean recognition score of 24.69 ($SD = 6.25$) and the control group 26.94 ($SD = 4.85$). Postman et al. (1948) reported a mean of 24.23 for the experimental group and 27.66 for the control group, a difference which was significant at the 1% level. The difference between the means of the experimental and the control groups in the present experiment was not significant, consequently no statement can be made concerning a facilitating or a depressing effect of recall upon recognition. If the replication were a success, as was indicated by the close agreement in the means of the experimental groups where the conditions were identical for the two experiments, it is reasonable to suppose that the significant difference which Postman et al. reported was, as they suggested, due to the further weakening of already weak traces during the 10 min. delay.

DISCUSSION

That correct recall may have a positive, and incorrect recall a negative, effect upon recognition, was established by Zangwill (1937, 1939) and Hanawalt (1937), and confirmed in the results of the present experiment. A recent survey (Hanawalt, 1958) has evaluated these earlier studies, and some other recent English experiments which pertain to the present problem, consequently discussion will be limited to the studies outlined in the introduction of this paper, the only ones which suggest an overall depressing effect of recall upon recognition.

The modified replication of the Postman et al. (1948) study clarified to some extent the relation of their study to the present one. When the time interval between learning and recognition was controlled, the evidence which Belbin (1950) and Kay and Skemp (1956) saw in it for the depressing effect of recall upon recognition disappeared. On the other hand, the failure of a facilitating effect of recall upon recognition to appear makes the evidence for it at the 8-min. interval in the present study more doubtful. It is quite obvious that there is neither a facilitating nor a depressing effect of recall present over all methods and material at this time interval. No evidence was discovered in the replication which would require a change in the interpretation of results by the original authors.

The reason for the apparent contradiction of the results of Belbin (1950) and of Kay and Skemp (1956) with those of the present experiment can scarcely be attributed to a different learning method since incidental learning was common to all three, nor to stimulus material since it was in each case meaningful and familiar. The chief differences between Belbin's experiment and the present one were in the difference in the learning situation, the recall activity, and in the type of recognition test, consequently these are the areas most likely to reveal an explanation of the difference in results. That the nature of the recall activity is important appears to be born out by the failure of Kay and Skemp to reproduce Belbin's results in their Exp. I on the identification of the picture as a whole. They accounted for this on the basis of an attitude change on the part of *Ss* because of the many-incident picture in contrast to Belbin's single-incident picture. This was doubtless a factor but there was another important difference: Belbin used "prompted by standard question" recall where Kay and Skemp used free written recall. Belbin's method of recall apparently facilitated the production of errors, and her recognition method the effective use of them. The importance of a detail on this type of test was exemplified by the responses of the 16 *Ss* of her nonrecall group to the changed caption on the picture. Even though they were directed to disregard it, 7 of the 16 failed to identify the picture, which reduced this half of the control group to a chance performance.

That the type of recognition test is important is borne out by the fact that Kay and Skemp did succeed in reproducing the Belbin effect with their 30-item questionnaire where they had failed on the picture as a whole. The method was essentially the same but *S* had 30 opportunities to say yes or no instead of one. Since the type of recognition was the outstanding difference in the design of Kay and Skemp's Exp. I and the present one, it is possible that this was an important factor in the divergence of results. In support of this possibility is the fact that the depressing effect of recall upon recognition appeared in neither the present experiment nor in the replication of the Postman et al. experiment, both of which used a multiple choice type of recognition test.

Another factor of importance in the difference in results between Kay and Skemp's Exp. I and the present one, is the difference in the percentage of correct recall: 24% compared to the present 45%. Since there appears to be no disagreement concerning the facilitating effect of correct recall upon recognition, the higher percentage of correct recall of the present experiment would boost the scores of the recall groups. The facilitating effect for this time interval (Table 2) was limited to the 11 frequently recalled items. Of course a high error score would also produce a depressing effect. Since the percentage of error of Kay and Skemp's experiment and the present one was approximately the same (21% to 20%) this may not have been too important. However, it must be remembered that 39% of the recall errors did not appear on the recognition blank of the present study and hence were probably ineffective. In the light of Belbin's results an erroneous recall appears to be much more effective on a recognition test of the type used by Kay and Skemp than on a multiple choice type, but possibly less effective in determining the strength of faint memory traces. Factors aside from the specific trace of the learning period, such as attitude, inference, etc. would seem to be more important in the former.

A further word should be said concerning the learning methods. Both Belbin and Kay and Skemp had *Ss* simply observe the pictures, while Postman et al. and the present experiment required a specific reaction to each item. Kay and Skemp reported some introspective evidence indicating that many associations were to the picture as a whole. It is possible that some details were not reacted to at all which would explain the low percentage of recall. A fleeting reaction to other items would tend to increase the error in recall and consequently, for this group, the error in recognition.

The above discussion of the divergent results concerns chiefly the 8-min. recognition period, since the present experiment is the only one which tested after several days. At this longer time interval there appears to be no doubt that recall has a facilitating effect under the conditions of the present experiment. The results are consistent with those of Hanawalt (1937, pp. 57–63) on memory for designs, where recall was found to have a facilitating effect for at least 8 wk. When learning is incidental, it is reasonable that the effect of recall should increase with delay in recognition, provided that there is more correct than incorrect recall, and the recognition test is one which favors the operation of traces of the learning period over general responses of the *S*. When recall and recognition follow immediately after learning, the traces are all relatively fresh so that traces strengthened by recall have less chance to show their strength. With the passing of time more of the incidental learning traces drop below functional strength for recognition, but many of the relatively stronger recall traces, whether correct or incorrect, are still above threshold value, as was inferred above in connection with Table 2.

Recognition, as Woodworth (1948, p. 81) pointed out for recall, is a problem to which *S* brings all of his "available responses," only some of which are based upon the traces of the learning period. Correct recall preceding recognition raises the score, and incorrect recall lowers it, but this is not the whole story. Even when the trace has been strengthened by correct recall, there are still erroneous recognitions, and correct recognitions (as many as one-third in the present results) when the trace has been weakened by erroneous recall. The *S*'s response on the recognition test is apparently influenced by subtle experimental factors, as suggested above, which bring variable general responses to bear upon the final decision.

SUMMARY

The effect of interpolated recall upon recognition was measured at different time intervals: immediately following recall and 48 or 52 hr. after recall. The learning was incidental in the form of a true-false test and recognition was measured by a multiple choice test. Interpolated recall produced some equivocal evidence for a facilitating effect of recall upon recognition when recognition followed immediately after recall; when recognition was delayed 48 or 52 hr., the facilitating effect was unequivocal. A test of the hypothesis that in recognition better known items have a depressing effect upon worse known items produced negative results. The effect of both erroneous and correct recall was found to increase with the lapse of time. The disagreement of the present results with some recent studies showing a depressing effect of recall upon recognition is discussed in the light of possible factors accounting for the differences. More research is needed before it will be possible to predict when recall will have a facilitating, a depressing, or no effect upon recognition.

REFERENCES

BELBIN, E. The influence of interpolated recall upon recognition. *Quart. J. exp. Psychol.*, 1950, **2**, 163–169.

HANAWALT, N. G. Memory trace for figures in recall and recognition. *Arch. Psychol., NY*, 1937, No. 216.

HANAWALT, N. G. Remembering. In G. S. SEWARD & J. P. Seward (Eds.), *Current psychological issues*. New York: Holt, 1958. Pp. 53–85.

KAY, H., & SKEMP, R. Different thresholds for recognition: Further experiments on interpolated recall and recognition. *Quart. J. exp. Psychol.*, 1956, **8**, 153–162.

POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *Amer. J. Psychol.*, 1948, **61**, 511–519.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1948.

ZANGWILL, O. L. An investigation of the relationship between the process of reproducing and recognizing simple figures with special reference to Koffka's trace theory. *Brit. J. Psychol.*, 1937, **27**, 250–276.

ZANGWILL, O. L. Some relations between reproducing and recognizing prose material. *Brit. J. Psychol.*, 1939, **29**, 370–382.

(Received July 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 368–371

THE RELATIVE EFFECTIVENESS OF POSITIVE AND NEGATIVE VERBAL REINFORCERS

AUSTIN JONES[1]

University of Pittsburgh

Buchwald (1959a, 1959b) and Buss and his associates (Buss, Braden, Orgel, & Buss, 1956; Buss & Buss, 1956; Buss, Wiener, & Buss, 1954) have studied acquisition and extinction as a function of three different combinations of verbal reinforcers—Right-Wrong, Nothing-Wrong, and Right-Nothing. It was uniformly found that acquisition is weaker under the Right-Nothing combination than under Nothing-Wrong and Right-Wrong. No significant differences have been found between the latter two. To account for the lesser efficacy of the Right-Nothing combination, Buss, Braden, Orgel, and Buss (1956) proposed that Nothing is a nonreinforcer and that Right is a relatively weak positive reinforcer, while Wrong is a relatively strong negative reinforcer. Buchwald (1959a) cites evidence in support of an alternative explanation, "that when *S* is exposed to either *RN* or to *NW*, *N* acquires a reinforcement value opposite in direction to that of the event with which it is combined, and that the value acquired by *N* in the *NW* combination exceeds that which it acquires in the *RN* combination" (p. 351). It should be noted that the two formulations are not mutually exclusive; it is conceivable that Right may be shown to be a weaker reinforcer than Wrong and also that simultaneously Nothing tends to acquire a value opposite to that of the word with which it is paired.

The purpose of the present study was to assist in the clarification of these issues by providing a direct test of the hypothesis that Right is a weaker reinforcer than Wrong. Each of the studies cited above provides only inferential evidence on the question, since the two words were always arranged in a combination, either with each other or with Nothing. In the present experiment, acquisition was studied as a function of "pure" positive and negative reinforcements. Subjects were given but a single training trial and were reinforced by the single utterance of the word Right or Wrong.

METHOD

Subjects.—In order to enhance the comparability of this experiment to those cited above, two different *S* samples were employed—the first a psychiatric group similar to that used by Buss, Braden, Orgel, and Buss (1956), the other an undergraduate college group similar to those used by Buchwald (1959a, 1959b). The psychiatric group consisted of 26 inpatients at a VA hospital, who served on a volunteer basis. Most of them were under 30 years, and had received 12 or more years of education. Patients were excluded who had received shock treatment within the preceding 3 weeks, who were disoriented, or who had organic diagnoses. The college group consisted of 34 undergraduate volunteers from the University of Pittsburgh.

Learning materials.—The task set for the *S*s was to learn the association of the nonsense syllable DAK with the number concept "four" in the Wisconsin Card Sorting Test (Grant, 1951). The WCST materials have been employed previously in the study by Buss and Buss (1956). In the present experiment, 64

---

[1] The author wishes to thank Arnold Buss and William Meyer for reading the manuscript, and H. Jean Wilkinson for assistance in the collection of data.

cards were used, with each color, form, and number of forms presented equally. A special training card was prepared which was of the same overall size and shape as the WCST cards. On it were drawn, in black ink on a white ground, four identical forms not represented anywhere in the WCST. These forms could be described approximately as verticle rectangles with a crescent-shaped notch in the upper end. There are no black and white figures in the WCST; thus, the only formal element shared by the training card and the WCST series was that of "fourness."

Procedure.—The *S*s were tested individually. They were seated across a desk from *E*, who informed them that this was "an experiment to see how well people can solve problems when they have very little information to go on." They were told they would be shown a series of cards, some of which are called DAK, the others VEC, and that they should try to sort them into the appropriate categories. Before being handed the deck of 64 cards, *S*s were shown the training card and asked to guess its identity. If the *S* said DAK, *E* replied "Right"; if *S* said VEC, *E* replied "Wrong." The deck of cards was then sorted without further communication between *S* and *E*.

TABLE 1

NUMBER OF *S*S SHOWING POSITIVE GRADIENTS (CONSISTENT WITH LEARNING THE CORRECT RESPONSE) AND ZERO OR NEGATIVE GRADIENTS, AFTER REINFORCEMENTS RIGHT AND WRONG

Reinf.	College Group $N = 34$			Psychiatric Group $N = 26$		
	Positive Gradient	Zero or Negative Gradient	Total	Positive Gradient	Zero or Negative Gradient	Total
Right	2	17	19	9	6	15
Wrong	9	6	15	3	8	11
Total	11	23	34	12	14	26

RESULTS

Learning of the correct solution—that DAK equals the number concept "four"—would be reflected in a generalization gradient showing a decreasing frequency of DAK responses over the categories 4, 3, 2, and 1 of the WCST. Failure to learn the correct solution would be indicated by flat or increasing gradients. An analysis of variance comparing the gradients associated with the two *S* groups and two reinforcers yielded no significant differences. The *S* variability was very great, as was expected in light of the unstructured, ambiguous quality of a task in which performance has been guided by a single reinforcement. A nonvariance type of analysis was then selected, in which each *S*'s performance was evaluated grossly for overall evidence of learning, regardless of the slope of the gradient. This analysis, utilizing a different type of score, is regarded as essentially independent of the unsatisfactory previous analysis. Table 1 shows the number of *S*s whose gradients were in the downward or "positive" direction indicative of correct learning—regardless of the magnitude of the trend—and the number whose gradients were zero or "negative." Thus, *S*s were categorized as having positive gradients whenever the frequency of DAK responses to the pooled four and three category exceeded that to the pooled two and one category. Fisher's exact probability test (Edwards, 1950) was used to assess the significance of the differential effectiveness of Right and Wrong in the two groups. The exact probability tests indicated: (*a*) in the college group, a significantly greater proportion of those receiving Wrong had positive generalization gradients than did those receiving Right ($P = .006$); (*b*) in the psychiatric group, the proportion of positive gradients among those receiving Right was not significantly greater than among those receiving Wrong ($P = .208$); (*c*) viewing the data from another direction, the psychiatric group has a significantly higher proportion of positive gradients under the Right condition than does the college group ($P = .006$), and (*d*) the proportion of positive gradients under the Wrong condition is not significantly higher in the college group than in the psychiatric group ($P = .208$). All probability values are given for two-tailed tests.

DISCUSSION

For the group of college *S*s, the results support the hypothesis that Right is a weaker reinforcer than Wrong. The group of *S*s receiving Wrong showed a proportion of "positive" gradients, i.e., those consistent with learning of the correct response, that was significantly greater than that of the *S*s receiving Right. Upon inspection of the absolute gradients of those receiving Right, it was apparent that, as a group, their gradients were not simply relatively flat, but that instead they moved in the upward or "negative" direction in a trend of approximately equal magnitude to the positive gradients of those receiving Wrong. This suggests that some learning had occurred, but of an incorrect concept in which DAK was associated with the numbers two and one.

Informal discussion with *S*s concerning their rationale for sorting of the WCST cards led to the following tentative explanation of the differential effectiveness of Right and Wrong for the college group. When an *S* responds to the single training card, he probably does so in accordance with an "hypothesis" or ideational response of some sort concerning the aspect of the stimulus card which denotes DAK-ness or VEC-ness. This initial hypothesis is almost certain to be incorrect because of the ambiguity of the task. If *S* says VEC, *E* says Wrong, thereby punishing not only the utterance of the incorrect syllable, but also the incorrect ideational response associated with it. The *S* thus turns to the series of 64 cards with his dominant and incorrect hypothesis having been punished, and with the correct hypothesis now somewhat closer to a position of dominance in his hierarchy of ideational responses. If, on the other hand, *S* happens to say DAK to the training card, *E* says Right and thereby reinforces not only the utterance of that syllable but also the incorrect hypothesis which preceded it. Thus, such an *S* begins to sort the cards with his initially dominant and incorrect hypothesis made still more dominant by *E*'s reinforcement. Consequently, he would be expected to be less "flexible" in his sorting, and more likely to perpetuate the type of error associated with his first hypothesis.

In general, it is suggested that Wrong will have a greater positive effect on learning than will Right when the dominant hypotheses concerning the discriminative stimuli are incorrect. An answer to the question of the differential effectiveness of positive and negative reinforcers per se would then await research in which the relative initial dominance of correct and of incorrect hypotheses is equated.

For the psychiatric group, the Buss, Braden, Orgel, and Buss (1956) hypothesis that Wrong is a stronger reinforcer than Right was not supported; as noted earlier the trend was in the opposite direction. The interpretation of this negative finding does not seem immediately clear.

SUMMARY

A direct test was attempted of the hypothesis that the word Wrong is a stronger reinforcer than the word Right. Twenty-six psychiatric inpatients and 34 college undergraduates learned the association of a nonsense syllable with the number concept "four." After single training trial in which *S*'s response was followed by *E* saying either "Right" or "Wrong," *S*s sorted 64 cards of the Wisconsin Card Sorting Test without further guidance from *E*. The data consisted of the number of *S*s whose generalization gradients showed decreasing frequencies of response from the pooled number category 4-3 to that of 2-1. In the college group such "positive" gradients, indicative of learning of the correct response, were significantly more often associated with the reinforcer Wrong than with Right. Such a significant difference did not appear in the data for the psychiatric group. A comparison between *S*

groups showed that Right was significantly more often associated with positive gradients in the psychiatric group than in the college group.

REFERENCES

BUCHWALD, A. M. Experimental alterations in the effectiveness of verbal reinforcement combinations. *J. exp. Psychol.*, 1959, **57**, 351–361. (a)

BUCHWALD, A. M. Extinction after acquisition under different verbal reinforcement combinations. *J. exp. Psychol.*, 1959, **57**, 43–48. (b)

BUSS, A. H., BRADEN, W., ORGEL, A., & BUSS, E. Acquisition and extinction with different verbal reinforcement combinations. *J. exp. Psychol.*, 1956, **52**, 288–295.

BUSS, A. H., & BUSS, E. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, **52**, 283–287.

BUSS, A. H., WIENER, M., & BUSS, E. Stimulus generalization as a function of verbal reinforcement combinations. *J. exp. Psychol.*, 1954, **48**, 433–436.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

GRANT, D. A. Perceptual versus analytical responses to the number concept of a Weigl-type sorting test. *J. exp. Psychol.*, 1951, **41**, 23–29.

(Received July 26, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 372–376

PATTERN MATCHING IN THE PRESENCE OF VISUAL NOISE[1]

MALCOLM D. ARNOULT
Texas Christian University

AND

CHARLES W. PRICE
University of Mississippi

During the past few years considerable research has been devoted to specifying the conditions affecting discrimination, recognition, and identification of shapes and patterns. Of particular interest is the situation in which a stimulus is to be matched with one which differs somewhat from it in appearance, because this situation could be considered a prototype of everyday perception. Completely identical objects are rarely seen, yet observers carry out fairly efficiently the processes of categorization and recognition despite the many differences which exist between objects belonging to the same class. Research is needed to identify the various ways in which visual stimuli may be distorted and the effects of distortion in relation to the kind of stimulus, the kind of perceptual task, and the conditions of observation.

Distortion has been used as an independent variable in experiments with both shapes and patterns, but the results obtained so far have not been easy to generalize from one experiment to another. Consequently, only the results obtained with distorted patterns will be discussed in relation to the present experiment.

French (1954) used patterns composed of small numbers of dots and introduced visual noise (distortion) by adding dots randomly to the pattern. He found that errors decreased as the number of dots in the basic pattern increased (up to nine) and increased as the number of noise dots increased (up to eight). Pollack (1955) investigated the effect of visual noise on the recognition of patterns constructed by the sequential arrangement of stimulus elements into "paths." In one of his conditions *S* was required to match a distorted path to one of a set of alternative paths. When the number of alternatives was four, for example, he found that a 50% distortion (noise level) of the stimulus path produced 20% errors in matching, corresponding to an increase of response uncertainty of slightly less than 1 bit.

Recently Hillix (1960) has reported an experiment in which the stimuli consisted of patterns of filled and unfilled cells of a 100 × 100 matrix. He varied the number of filled cells from 10% to 50% and the amount of distortion from 10% to 40%. He found that the number of errors made in matching distorted patterns to their prototypes could be predicted ($r^2 = .88$) from a similarity index based on the number of filled cells common to the two stimuli. Likewise, response latency could be predicted on the same basis ($r^2 = .94$).

The present experiment was, in general, very similar to the one by Hillix, although the specific conditions all differed. Cells were filled with dots rather than squares, the number of filled cells was determined on a probability basis ($P = .50$), distortion was introduced by adding or removing filled cells rather than by displacing them, and the conditions of stimulus presentation differed in a number of details. Nevertheless, there was considerable similarity between the tasks, and it is of interest to determine whether the data can be described by the same empirical function.

Method

Stimuli.—A number of *prototype* patterns was constructed, and for each of these a number of *test patterns* was constructed by introducing a specified amount of randomly-determined variation into the prototype pattern. The prototypes were constructed on 8 × 8 matrices, with dots being placed in the 64 cells with a probability of .50. The actual number of dots on the completed prototypes varied from 24 to 36, with a median of 29.5. Four prototypes were assigned to each of four conditions of visual noise, and 7 test patterns constructed for each prototype, making a total of 112 test patterns in the experiment.

---

[1] This report is based in part on a thesis submitted by the second author to the Graduate School of the University of Mississippi in partial fulfillment of the requirements for the degree of Master of Arts in January 1960.

Test patterns were constructed by distorting the prototypes in the following way: a randomly selected cell of the matrix was changed to the opposite state; i.e., if it was vacant, a dot was added, and if it had a dot, the dot was deleted. The number of cells chosen for alteration was determined by the noise level desired. In the 0% noise condition no cells were changed; each test pattern was identical to its prototype. The 5% noise level was approximated by changing 3 of the 64 cells. Similarly, 10% noise involved changing 6 cells, and 20% noise involved changing 13 cells. Thus, each test pattern at a given noise level was equivalent to each other test pattern at that level but differed in terms of the particular cells which had been changed.

The test patterns were made by photographing a matrix of illuminated circles. The resulting negatives were bound into 2 × 2 in. slides for use in the projector. The prototype patterns were cut out of heavy poster board and backed by black paper. The dot pattern was centered on a 10 × 10 in. field, which was the size of the field on which the test patterns would be projected. Figure 1 is an example of the test situation as viewed by *S*.

Apparatus.—The test patterns were projected by an ordinary slide projector equipped with an aluminum leaf-type shutter controlled by a rotary solenoid actuated by a Hunter timer. Exposure time was 0.7 sec.; interexposure time was controlled by *E* and was about 5 sec. in duration.

The projection screen was a 37 × 38 in. Masonite board painted flat black on *S*'s side. The four prototype patterns for a given noise level were attached at the corners of the screen. In the center was a 10 × 10 in. translucent screen on which the test patterns were projected from the rear. The room was brightly illuminated by overhead fluorescent fixtures, and the contrast in the projected pattern was approximately equal to that of the prototype patterns.

Procedure.—The *S*s were seated at an average distance of 11.5 ft. from the screen and provided with answer sheets. For each trial the answer sheet provided four boxes corresponding to the positions of the four prototype patterns. The following instructions were read:

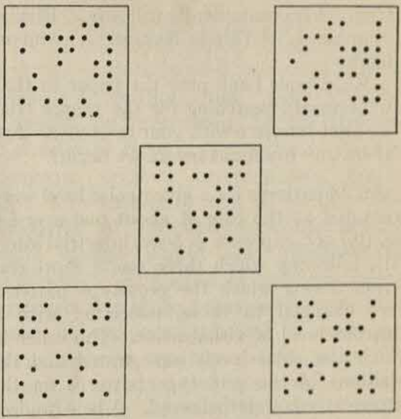


FIG. 1. The matching task as viewed by *S*. (The test pattern appeared in the center for 0.7 sec. and was to be matched with one of the four surrounding prototypes. The pattern illustrated is a 20% distortion of the lower left prototype.)

> The purpose of this experiment is to investigate how accurately and rapidly people can identify complex patterns under various conditions. All of the patterns you will see will be of the same general type as the four which you now see in front of you. I will flash a pattern on this screen in the center; your task will be to determine which of these other four patterns it most resembles. In some cases the resemblance will be quite obvious. In other cases your decision will be more difficult. But in every case the pattern which is flashed on the screen will resemble one of these four "prototypes" more than it resembles the other three. Is that clear?
>
> The test pattern will remain on the screen only a short time, so you must pay close attention if you are to identify it correctly. As soon as the pattern disappears, mark your decision on your answer sheet in one of the four boxes provided for each trial. If the test pattern most resembles the upper-left prototype, put an "X" in the upper-left box; if it most resembles the lower-right prototype, put your "X" in the lower-right box, and so forth. Is that clear?
>
> Just before each pattern is flashed on the screen I will give you a "Ready" signal by calling out the number of the

> trial. For example, I will say "This is Number 1," "This is Number 2," and so forth.
>
> Keep your hand near the paper so that unnecessary searching for the proper trial will not interfere with your responses. Are there any questions before we begin?

All 28 patterns for a given noise level were presented at the rate of about one every 6 sec. (0.7-sec. exposure + 5-sec. intertrial interval), following which there was a short rest period during which the prototype patterns were changed to those associated with a different level of visual noise. The order in which the noise levels were tested and the positions of the prototype patterns on the screen were counterbalanced. A new random order of the 28 test patterns within a noise level was used with every group of *S*s.

Subjects.—The *S*s were 80 college students about equally divided between men and women. They were unsystematically divided into 10 groups of 8 *S*s each, and data were collected from all members of a group simultaneously.

RESULTS AND DISCUSSION

The data were analyzed both in terms of the percentage of errors at each noise level and in terms of amount of information transmitted as a function of amount of visual noise, or distortion (Attneave, 1959). Figure 2 shows that both measures yielded functions which do not depart significantly from linearity over the range of noise levels used. At the highest noise level *S*s were making slightly over 50% errors and were transmitting only 0.20 bits of information of the 2.00 bits inherent in the four response categories.

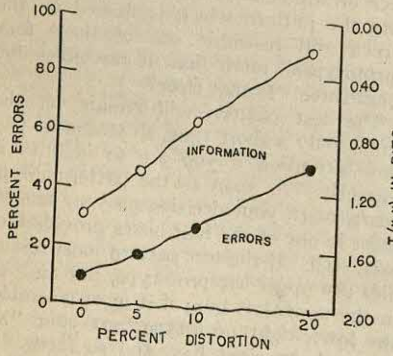


FIG. 2. Percentage errors and information transmitted as a function of the percentage distortion (noise level) of the prototype patterns.

It is interesting to compare these curves with the data reported by Pollack (1955) for distorted dot patterns arranged to form "paths." He found errors to be a positively accelerated function of amount of visual noise, with little or no increase in errors below a noise level of 20%. Although the methods of producing noise are not directly comparable, it is notable that there are only slight signs of positive acceleration in the error function obtained in the present experiment. It is also notable that there were 10% errors committed at the 0% noise level in this task and only about 1% in Pollack's task, suggesting that the present task was considerably more difficult. The number of pattern elements was roughly the same in the two studies (28 vs. a median of 29.5), and the total number of matrix cells was much greater in Pollack's experiment (324 vs. 64), but the constraint he imposed in requiring the elements to form a path apparently reduced the total amount of information in the patterns to a value less than in the dot patterns used here.

The present task was also more difficult than that used by Hillix (1960). His 50% Fill- 10% Distortion condition and 50% Fill- 20% Distortion condition are comparable, respectively, to the 10% and 20% distortion conditions we used. His *S*s committed about 25% errors in both conditions whereas our *S*s committed 27% and 50% errors. The difference is probably attributable to three factors: (*a*) Hillix' *S*s had to choose from among three alternatives, whereas our *S*s had four; (*b*) different methods were used to produce distortion; and (*c*) his stimuli were exposed until *S* responded (about 9 sec.), whereas our stimuli were exposed for only 0.7 sec. Despite the differences in overall level of performance in the two experiments, *S*'s responses may have been determined by the same sort of factors in both cases. Hillix was able to show that he could obtain an r^2 of .88 between the number of correct responses and a measure of "similarity" based, essentially, on the number of filled cells common to the standard and the correct alternative. Consequently, an attempt was made to apply a similar measure of similarity to the present stimuli. It was not possible to use exactly the same index used by Hillix because the number of filled cells in common with the standard was not necessarily the same for each of the seven alternatives constructed. Instead, the proportion of filled cells in common was computed for each alternative and averaged across all seven. This procedure resulted in 16 values for the similarity index, one for each of the four prototypes (or standards) at each of the four noise levels. Correlating the values of this index with the number of correct choices for each prototype at each noise level yielded an $r^2 = .94$, with the equation for the line of best fit being $C' = 1.977I - 1.070$, where C' is the predicted number correct and I is the index of similarity.

Thus, for both Hillix' experiment and the present one, it is possible to make quite accurate predictions of the number of correct responses as a function of the amount of distortion introduced into the pattern. To be most useful, however, the index of similarity should predict not only that an error will be made but which incorrect alternative will be chosen. Every possible pair of patterns should have some probability of being chosen as a consequence of the number of filled cells in common. At each level of distortion in the present experiment there were 84 possible pairs composed of test patterns and incorrect prototypes. The index of similarity for each pair was correlated with the number of times each pair was chosen in the experiment. The correlations for the various noise levels fell in the range from .25 to .35. These were statistically significant, but were not large enough to provide a basis for accurate predictions. It may be that precise prediction of pairings of these patterns depends upon complex subpatterns of dots as well as upon the locations of individual dots. For example, it would be interesting to get indexes of similarity based upon pairs of dots in common, triplets in common, quadruplets in common, and so forth. The similarity of subpatterns composed of white spaces would also possibly be important. The kind of analysis being proposed would be practicable only if it could be automatized and carried out by a large computer. If it paid off, though, it would provide a possible basis for predicting errors of matching made with almost any kind of complex patterns, in that the patterns could be quantized as matrices of filled and unfilled cells, and the index of similarity used to predict the probability of matching errors made by a human observer.

In analyzing the incorrect responses it was observed that these errors appeared to be nonrandomly distributed among the prototypes and that this tendency toward nonrandomness appeared to increase with increasing distortion of the test patterns. This observation was tested by computing the value of χ^2 for the distribution of errors among incorrect alternatives for each prototype at each level of distortion. The probabilities associated with each of the resulting 16 χ^2's are shown in Table 1. It can be seen that in 8 of the 16 cases χ^2 was significant beyond the .05 level, and

TABLE 1

Probabilities Associated with χ^2 Tests of the Hypothesis that Errors are Distributed Randomly among the Incorrect Alternatives

Prototype	Noise Level			
	0%	5%	10%	20%
A	*	<.01	*	<.01
B	*	*	*	<.01
C	*	<.01	<.01	<.01
D	<.05	*	<.01	*

*$P > .05$.

that both the number of significant χ^2's and their level of significance increased with an increase in the amount of distortion. It would appear, then, that one of the effects of increasing distortion was to cause the test patterns to become more similar to some of the incorrect prototypes than to others.

SUMMARY

Prototype patterns of dots were constructed by placing dots with a probability of .50 in the cells of an 8 × 8 matrix. Distorted versions of these prototype patterns were then constructed by randomly removing or adding dots in 0%, 5%, 10%, or 20% of the cells in the matrix. Four prototype patterns were used at each noise level, and seven randomly-differing test patterns were constructed for each prototype. Each test pattern was presented for 0.7 sec., and *S*'s ($N = 80$ college students) task was to match it to the prototype which it most resembled.

The major findings were:

1. Errors increased from 10% at 0% distortion to 50% at the 20% distortion level. At the same time, information transmitted decreased from 1.38 bits to 0.20 bits. Both relationships were linear.

2. The increase in errors could be accounted for ($r^2 = .94$) by an index of similarity based on the number of dots in common between the test patterns and the prototype at each noise level. The index did not, however, account satisfactorily for the distribution of errors among the incorrect alternatives.

3. The distribution of errors among incorrect prototypes was increasingly nonrandom as distortion increased.

REFERENCES

ATTNEAVE, F. *Applications of information theory to psychology.* New York: Holt, 1959.

FRENCH, R. S. Pattern recognition in the presence of visual noise. *J. exp. Psychol.,* 1954, **47**, 27–31.

HILLIX, W. A. Visual pattern identification as a function of fill and distortion. *J. exp. Psychol.*, 1960. **59**, 192–127.

POLLACK, I. Visual noise filtering by human operators: I. Sequentially coded information. *USAF Operat. Applications Lab. tech. Rep.*, 1955, No. AFCRC-TR-54-57.

(Received July 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 377–384

DEPRIVATION AND REINFORCEMENT[1]

GEORGE COLLIER AND FRANK N. WILLIS

University of Missouri

On the basis of a series of studies (Collier & Myers, 1961; Collier & Siskel, 1959), two clusters of events have been isolated which control momentary rate of responding. The first cluster is the olfactory, gustatory, kinaesthetic, and other sensory consequences of the reinforcing substance. The basic relation here is that rate is a linear increasing function of log intensity of stimulation. The second cluster is the momentary postingestive load. Rate of responding is a decreasing function of concentration. The present study is concerned with the effect on these functions of the nutritive state of the animal.

METHOD

Subjects.—The 172 *S*s were naive male rats, 90 to 120 days old, of the Sprague-Dawley strain obtained from the Holtzman Company. Twenty-eight *S*s were lost because they either failed to drink during magazine training or failed to make more than 10 responses per bar press (BP) session; 23 of the losses occurred in the low-deprivation condition and 12 of these were lost in the 4% condition. This differential loss biases the sample.

Apparatus.—Four Skinner boxes that provided liquid reinforcement were used. The magazines were 12-in. diameter, covered, aluminum plates with 72 .4-ml. cups around the periphery. The magazines were loaded by means of a Cornwall automatic pipetting device. A more detailed description of the apparatus is given in Collier and Myers (1961). The solutions used were prepared from commercial sugar 24 hr. before use and kept at room temperature no longer than 48 hr.

Procedure.—Three concentrations of sucrose (4%, 16%, and 64%), two volumes of reinforcement (.1 and .3 ml.), two fixed interval, interreinforcement intervals (1 and 4 min.), and two levels of deprivation (3 hr. [L] and 23 hr. [H]) were combined factorially. Six *S*s were in each cell. Three hours was chosen as the low deprivation condition to minimize gastric load. Experimental sessions were 30 min. in duration. Seven days of table training involving handling deprivation adaptation, and drinking of the appropriate test solutions, were followed by 4 days of magazine training, 7 days of bar press training, and 3 days of extinction.

On the 3-hr. deprivation schedule, the food was removed 3 hr. before the experimental session and replaced immediately following the session. On the 23-hr. schedule, *S*s were fed for 1 hr. immediately following the experimental session. The differential amounts of sugar ingested under each of the Concentration × Volume × Interval combinations were not compensated for in the daily diet.

Magazine training consisted of 30 presentations per session of the appropriate volume and concentration on a 1-min. variable interval (VI) schedule. The number of presentations consumed was recorded and any *S* which failed to take at least 25 of the 30 possible on each of the last 2 days of magazine training was discarded.

During extinction sessions, the magazine was present but did not operate.

The experiment was run in replications of 24 *S*s. With the exception of volume, conditions were counterbalanced across cages, boxes, and time of running. The *S*s that received .3 ml. reinforcement were started when about one-half of those with .1 ml. reinforcement were completed.

RESULTS

Figure 1 presents the cumulative BP curves showing the within session course of responding. Figure 2 presents the number of responses in the first 5 min. An analysis of the average

---

[1] This investigation was supported by Research Grant M-3328 from the National Institute of Mental Health, of the National Institutes of Health, Bethesda, Maryland. A preliminary report of these data was given in a paper entitled "Regulation and reinforcement," Midwestern Psychological Association, 1960, St. Louis, Missouri.

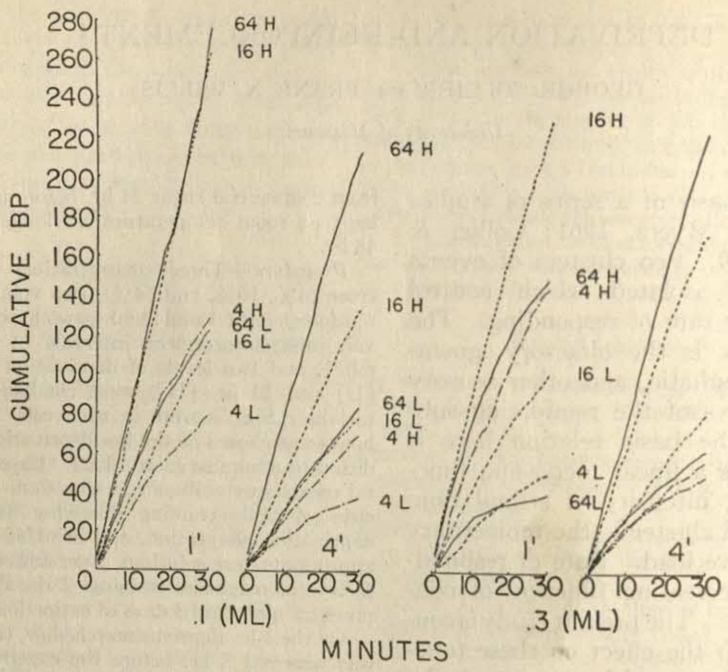


FIG. 1. Cumulative number of BP as a function of concentration (4%, 16%, 64%), volume (.1 and .3 ml.), interval (1 and 4 min.), and deprivation (L and H).

of the last 2 days (6 and 7) of reinforced BP for concentration, volume, interreinforcement interval, deprivation, and minute of the session is presented in Table 1. Figure 3 presents the paralle datal for Day 1 of extinction. An analysis of these data for C, V, I, D, and extinction sessions, is summarized in Table 2. Initial and terminal rate will be examined separately.

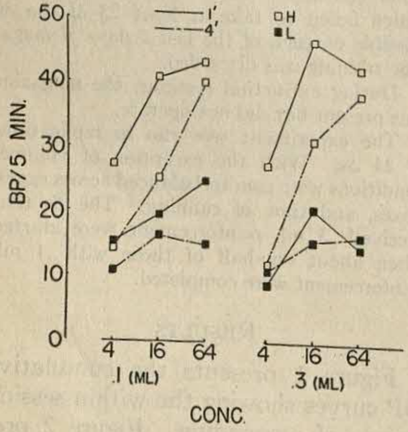


FIG. 2. Total number of BP in first 5 min. as a function of concentration (4%, 16%, 64%), volume (.1 and .3 ml.), interval (1 and 4 min.), and deprivation (L and H).

Initial rate.—Examination of the initial slopes of the curves in Fig. 1 shows that initial rate is an increasing function of concentration. Comparison of Fig. 3 with Fig. 2 further shows that the linear relation between rate and log concentration is obscured by the decremental effects of even a few reinforcements and their consummatory and postingestive consequents.

Initial rate is an increasing function of volume as shown in Fig. 2. A similar relation is shown in the extinction data of Fig. 3, where the consummatory and postingestive effects are absent. Volume and concentration interact (Fig. 2 and 4) so that the slope of the initial rate-concentration function is greater at higher volumes. Figure 4 presents the volume-concen-

TABLE 1

ANALYSIS OF VARIANCE OF THE AVERAGE NUMBER OF BP PER MINUTE OVER DAYS 6 AND 7

Source	*df*	*MS*	*F*
Between *S*s	143	983.9	
Concentration (C)	2	10,165.3	26.82**
Volume (V)	1	2,030.9	5.36*
Interval (I)	1	6,283.4	16.57**
Deprivation (D)	1	43,029.2	113.51**
C ×V	2	1,556.7	4.11*
C ×I	2	1,598.5	4.22*
C ×D	2	3,816.2	10.07**
V ×I	1	3,104.2	8.19**
V ×D	1	1.5	.00
I ×D	1	1,645.3	4.34*
C ×V ×I	2	884.2	2.33
C ×V ×D	2	20.8	.05
C ×I ×D	2	744.8	1.96
V ×I ×D	1	59.3	.16
C ×V ×I ×D	2	743.8	1.96
Error (b)	120	379.1	
Within *S*s	4176	13.8	
Minutes (M)	29	183.9	15.58**
C ×M	58	17.6	1.51**
V ×M	29	30.2	2.56**
I ×M	29	29.8	2.52**
D ×M	29	17.3	1.46
C ×V ×M	58	17.4	1.49
C ×I ×M	58	14.9	1.28
C ×D ×M	58	15.3	1.31
V ×I ×M	29	35.1	2.98**
V ×D ×M	29	20.1	1.70*
I ×D ×M	29	8.7	.74
C ×V ×I ×M	58	18.8	1.61**
C ×V ×D ×M	58	8.7	.75
C ×I ×D ×M	58	9.3	.79
V ×I ×D ×M	29	14.2	1.21
C ×V ×I ×D ×M	28	20.7	1.77**
Error (w)	3480	11.6	

* $P \leq .05$.
** $P \leq .01$.

tration extinction data collapsed over interval. In the present experiment at the lowest concentration the volumes used do not produce very different initial rates. Other data obtained in this laboratory indicate that this difference at the near threshold concentration is a function of the kind (e.g., thirst or hunger) and degree of deprivation.

Initial rate is a decreasing function of interreinforcement interval. This relation is preserved in the number of extinction responses of Day 1 of extinction, but on Days 2 and 3 the 4-min. interval groups show greater resistance to extinction (Table 3). While the interactions involved do not quite reach significance, it appears that the relation (Fig. 3 and Table 3) between interval and number of extinction responses in the first session is a complex one. The superiority in number of responses appears to hold only at small volumes and to be increased by deprivation (Fig. 3). The slope of the initial rate-concentration function is a decreasing function of interval (Fig. 1). This relation is very rapidly contaminated by consummatory and postingestive effects. As can be seen in Fig. 3, the greater resistance to extinction of the long-interval groups at large volumes confounds this relation in extinction, with the large-volume long-interval groups showing the greatest resistance to extinction. Neither the present study nor the Collier and Myers (1961) study yields any evidence of an interaction of volume and interval for initial rate, although there is a substantial one over the course of a session as will be noted later.

Initial rate is an increasing function of deprivation. Of greater novelty and more interest are its interactions.

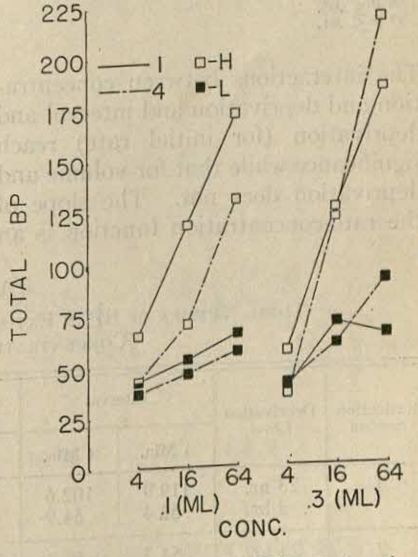


FIG. 3. Total number of BP in first extinction session as a function of concentration (4%, 16%, 64%), volume (.1 and .3 ml.), interval (1 and 4 min.), and deprivation (L and H).

TABLE 2

ANALYSIS OF VARIANCE OF TOTAL NUMBER OF BP IN THE THREE EXTINCTION SESSIONS

Source	*df*	*MS*	*F*
Between *S*s	143	4,789.7	
Volume (V)	1	6,896.0	2.35
Concentration (C)	2	68,377.3	23.38**
Interval (I)	1	46.7	.02
Deprivation (D)	1	106,220.1	36.32**
V × C	2	4,702.7	1.60
V × I	1	8,497.8	2.90
V × D	1	327.3	.11
C × I	2	1,974.9	.68
C × D	2	25,303.6	8.65*
I × D	1	1,095.7	.37
V × C × D	2	2,622.0	.90
V × C × I	2	1,714.8	.59
V × I × D	1	690.1	.24
C × I × D	2	409.7	.14
V × C × I × D	2	699.6	.24
Error (b)	120	2,919.6	
Within *S*s	288	1,440.9	
Extinction Sessions (E)	2	104,480.7	300.59**
V × E	2	2,134.9	6.14**
C × E	4	11,370.6	32.71**
D × E	2	17,974.7	51.71**
I × E	2	2,428.5	6.98**
V × D × E	4	615.9	1.77
V × I × E	2	758.1	2.18
V × D × E	2	635.7	1.82
C × I × E	4	319.4	.92
C × D × E	4	4,706.8	13.54**
I × D × E	2	899.6	2.58
V × C × I × E	4	428.6	1.23
V × C × D × E	4	179.9	.52
V × D × I × E	2	674.2	1.93
C × I × D × E	4	162.3	.47
V × D × I × D × E	4	114.1	.33
Error (w)	240	347.6	

* $P \leq .05$.
** $P \leqq .01$.

The interactions between concentration and deprivation and interval and deprivation (for initial rate) reach significance while that for volume and deprivation does not. The slope of the rate-concentration function is an increasing function of deprivation (Fig. 1, 2, 3, and 4). The course of this relation over extinction sessions is shown in Table 3. Table 3 further shows that the rate-volume functions for high and low deprivation are essentially parallel.[2] The effect of increasing deprivation on the rate-interval relation is to increase the slope of the rate-interval function.

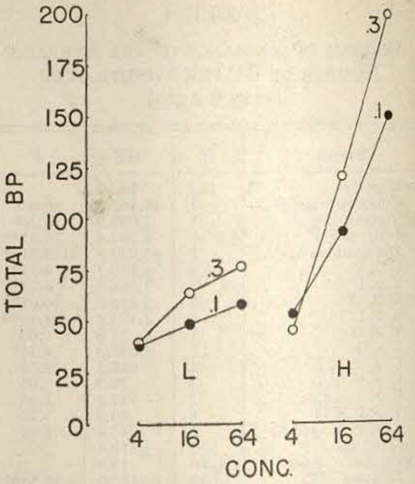


FIG. 4. Total number of BP in first extinction session as a function of concentration (4%, 16%, 64%), volume (.1 and .3 ml.), and deprivation (L and H).

TABLE 3

TOTAL NUMBER OF BP IN EXTINCTION AS A FUNCTION OF INTERVAL, CONCENTRATION, AND VOLUME

Extinction Session	Deprivation Level	Interval		Concentration			Volume	
		1 Min.	4 Min.	4%	16%	64%	.1 ml.	.3 ml.
1	23 hr.	119.9	102.6	49.3	107.7	176.2	99.3	122.6
	3 hr.	54.4	54.9	38.9	56.2	68.7	49.4	59.6
2	23 hr.	54.7	57.5	33.0	52.6	82.3	54.0	57.8
	3 hr.	31.1	33.9	25.5	32.2	39.6	29.6	35.7
3	23 hr.	34.6	41.6	19.8	40.8	53.5	37.0	39.0
	3 hr.	20.1	28.3	21.4	22.3	28.7	22.7	25.1

---

[2] This finding has been verified in our laboratory in a recent study employing a wider range of volumes and using water for reinforcement.

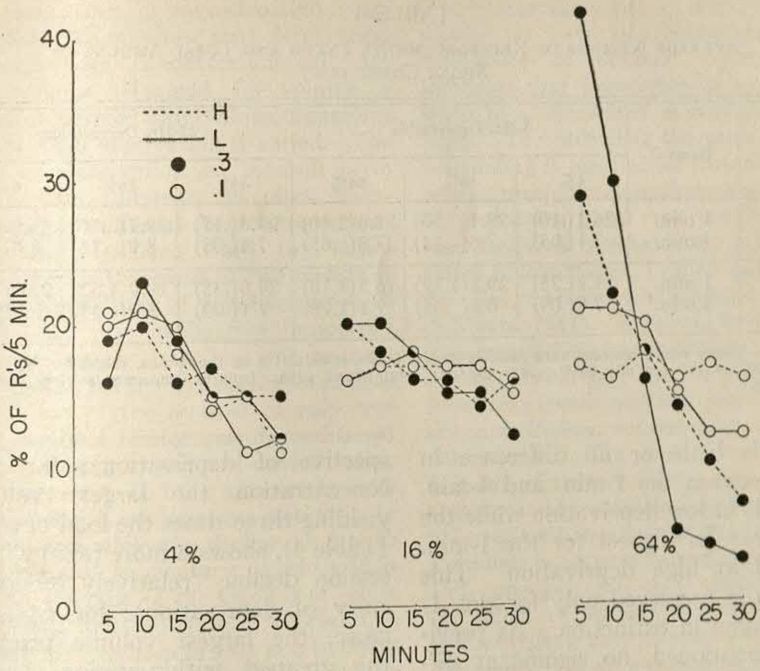


FIG. 5. Number of BP in each 5 min. of reinforced BP as a function of concentration (4%, 16%, 64%), volume (.1 and .3 ml.), and deprivation (L and H) for 1-min. FI.

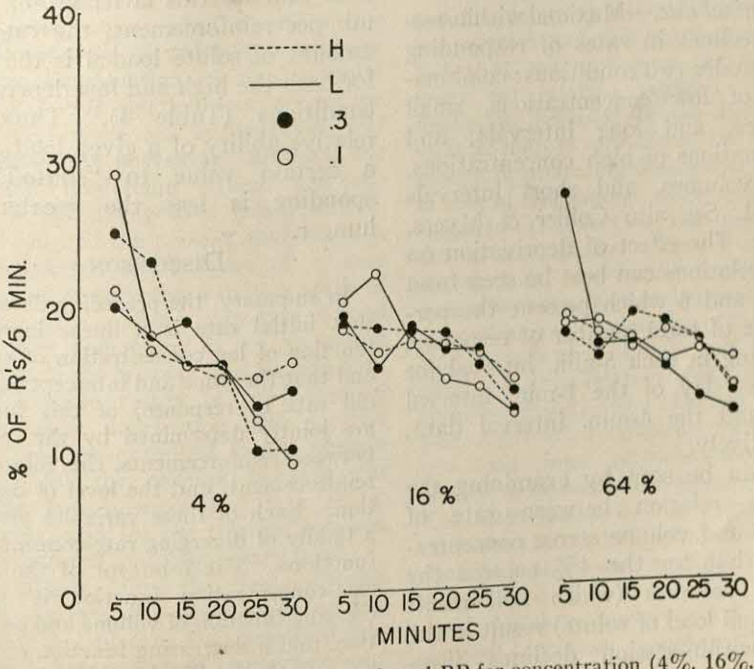


FIG. 6. Number of BP in each 5 min. of reinforced BP for concentration (4%, 16%, 64%), volume (.1 and .3 ml.), and deprivation (L and H) for 4-min. interval.

TABLE 4

AVERAGE NUMBER OF REINFORCEMENTS TAKEN AND TOTAL AMOUNT OF SUGAR CONSUMED

Volume	Interval	3-Hr. Deprivation			23-Hr. Deprivation		
		4%	16%	64%	4%	16%	64%
.1 ml.	1 min.	26.1(.10)	29.2(.50)	28.6(2.40)	29.3(.12)	29.7(.50)	29.9(2.51)
	4 min.	7.1(.03)	7.6(.13)	7.8(.65)	7.4(.03)	8.0(.14)	8.0(.67)
.3 ml.	1 min.	21.2(.25)	29.8(1.52)	16.3(4.10)	29.0(.35)	29.9(1.52)	29.0(7.29)
	4 min.	7.8(.09)	6.8(.35)	7.1(1.78)	7.1(.09)	8.0(.41)	8.0(2.00)

Note.—Thirty reinforcements were possible on the 1-min. schedule; 8 on the 4-min. schedule. 4% = .040 gm/ml; 16% = .170 gm/ml; 64% = .838 gm/ml; total grams sugar (solute) consumed/session shown in parentheses.

There is little or no difference in rates between the 1-min. and 4-min. intervals at low deprivation while the initial rate is highest for the 1-min. interval at high deprivation. This relation is preserved only for the .1-ml. volume in extinction. As previously mentioned, no significant differences in rate of responding for the two intervals occur at the .3-ml. volume in extinction.

Terminal rate.—Maximal within-session declines in rates of responding occur under two conditions: combinations of low concentrations, small volumes, and long intervals; and combinations of high concentrations, large volumes, and short intervals (Fig. 1. See also Collier & Myers, 1961). The effect of deprivation on these relations can best be seen from Fig. 5 and 6 which present the percentage of total number of responses occurring in each 5-min. interval for the last day of the 1-min. interval data and the 4-min. interval data, respectively.

It can be seen by examining the shifting relation between rate of decline and volume across concentrations, that for the 4% sucrose the largest volumes (which still result in a small load of solute) result in the least within-session declines (irrespective of deprivation); for 16% concentration the largest volume, yielding three times the load of solute (Table 4), shows a more rapid within-session decline (relatively independently of deprivation); for 64% sucrose, the largest volume produces the greatest within-session decline, and the low deprivation, the most rapid decline. It is interesting to note that for this latter group, at .1 ml. per reinforcement, the rate and amount of solute loaded is the same for both the high and low deprivation conditions (Table 4). Thus, the relative ability of a given load above a certain value to "shutoff" responding is less the greater the hunger.

DISCUSSION

In summary, the preceding data show that initial rate is a linear increasing function of log concentration of sucrose and that the slope and intercept (threshold rate of response) of this function are jointly determined by the interval between reinforcements, the volume per reinforcement, and the level of deprivation. Each of these variables produces a family of diverging rate-concentration functions. The intercept of the initial rate-concentration function is an increasing function of volume and deprivation, and a decreasing function of interval. Both the BP data and the extinction data (free of ingestive and postingestive effects) show that deprivation interacts with concentration but not with volume. Instead, for volume, a series of parallel rate-volume functions results when deprivation is varied. The effect of deprivation on interval is to increase the difference in rates of responding at different interreinforcement intervals. Declines from the initial rate occur as functions of two classes of events, the number of ingestive responses made and the colligative properties (osmotic pressure, etc.) of the momentary gastric load. The decline in rate as a function of the number of responses (i.e., minimal reinforcement conditions) is a decreasing function of concentration and appears to be little affected by volume (amount of ingestive activity) or deprivation while the decline in rate as a function of momentary gastric load is a joint function of the amount of solute consumed and the state of deprivation. Differential declines in rate as a function of these variables are responsible for the nonmonotonicity of all but the initial, rate-concentration functions. Variations in the character of a reinforcement are reflected in variations in the performance of the behavior of which it is a consequent. Examination of the temporal course of this relation has led to the view that there are at least two loci for these rate governing variables, the proximal reinforcing stimuli and the immediate postingestive concentration. The results of the present study show a to-be-expected third locus of a rate governing event—the nutritive state of the organism. Momentary rate of food reinforced responding is an increasing function of the hours of food deprivation. Two possible points in the ingestive chain where deprivation may exercise its effects are examined in the present study: (*a*) the relation between the sensory consequences and rate of responding, and (*b*) the relation between rate of shutoff and momentary load.

Considering, first, the effects of deprivation on rate of shutoff, there are two sets of variables to be considered, the ingestive responses and the gastric load. The occurrence of a within-sessions decline in rate of responding under conditions of minimal reinforcement indicates that occurrence of a response leads to a decrement in response likelihood. The fact that the same level of responding is maintained from session to session precludes an explanation in terms of extinction, and the small gastric load involved precludes an account in terms of satiation. It may perhaps be denoted as response habituation (Collier & Myers, 1961). As would be expected, it is insensitive to deprivation. Rate of shutoff, excluding the minimal reinforcement conditions, is a joint function of concentration, volume consumed, and deprivation. The interaction of the first two of these variables suggests that tonicity is the variable involved. That volume itself is not sufficient to produce a decline is apparent from the 4% and 16% groups. At high concentrations the effect of volume is greater. The volume effect may simply reflect local dilution. This view is supported by the contrast between the 16% and 64% groups.

Sixteen percent sucrose is close to tonicity (approximately 9% in the rat for sucrose) and this is reflected in the fact that neither volume per reinforcement nor interval between reinforcements, as reflected in the load taken on in 20 min., is very effective in the 16% groups in producing differential declines in rate of responding. How deprivation exercises its effect is not clear from the present study. Two alternatives are that (*a*) deprivation modifies the proportionality relation between rate of responding and the tonicity of the gastric load, or (*b*) that it does not have an effect on this relation and the effect observed results from the fact that at 3-hr. deprivation sufficient residual gastric load from previous feedings is present to interact with the current load. The latter conclusion is suggested by the results of the 16% group which again show little effect as the result of different deprivations.

Of the three sensory parameters examined (volume, concentration, and interval) deprivation appears to interact with only the latter two. The slopes of the initial rate-concentration and initial rate-interval functions are increasing functions of deprivation while the slope of the initial rate-volume function is not.

The effects of deprivation and volume are additive, while those of concentration and deprivation and interval and deprivation are multiplicative. Thus it appears that deprivation exerts its influence on the rate of ingestion via an effect on the relation between the intensity of the reinforcing stimulus and the rate of performance of the reinforced response. Volume does not change the intensity of stimulation, and therefore, changes in deprivation should, on this view, result in a family of parallel rate-volume curves. At small volumes, where dilution may play a major role, volume and concentration interact in a fashion similar to the intensity-area functions (Collier & Myers, 1961). An account of the interval-deprivation interaction is not obvious. The times involved are too long for summation in the ordinary sense. If it is assumed that adaptation is proportional to intensity, then the effect observed could result from the greater recovery from adaptation at the long intervals.

An attractive hypothesis has been that deprivation lowers the absolute gustatory thresholds (e.g., Campbell & Sheffield, 1953; Hebb, 1955). The diverging rate-concentration functions having a common origin obtained in the present study do not support this conjecture. If rate of responding is equated with sensitivity, the present data are consistent with the hypothesis that differential gustatory thresholds are modified by deprivation; but it seems more reasonable to assume that rate and sensitivity are not equivalent and that the proportionality relation between intensity of stimulation and rate of responding is modified by deprivation. Such an assumption would account for such preference threshold data as those of Campbell (1958). In this study, which investigated the percentage of times that *S* will drink more of one concentration when compared with a second concentration or with water, Campbell found lower absolute and differential preference thresholds with increasing hours of deprivation. Although these results involved both sensory and postingestive factors, it seems that they are mainly the result of differential initial rates of responding and of differential rates of shutoff for the pairs of solutions compared and thus the result of changing proportionality between rate and intensity rather than changing thresholds.

SUMMARY

The relations between initial and terminal rate of bar pressing and concentration, volume per reinforcement, and interval between reinforcements were explored as functions of deprivation. The slopes of the initial rate-concentration and initial rate-interval functions were functions of deprivation while the initial rate-volume function was not. The rate of "shutoff" was most sensitive to deprivation at large concentrations, large volumes, and short intervals. Three independent loci of events controlling food ingestive behavior were suggested: the proximal reinforcing stimuli, the momentary postingestive load, and the nutritive state.

REFERENCES

CAMPBELL, B. A. Absolute and relative sucrose preference thresholds for hungry and satiated rats. *J. comp. physiol. Psychol.*, 1958, **51**, 795–800.

CAMPBELL, B. A., & SHEFFIELD, F. D. Relation of random activity to food deprivation. *J. comp. physiol. Psychol.*, 1953, **46**, 320–322.

COLLIER, G., & MYERS, L. The loci of reinforcement. *J. exp. Psychol.*, 1961, **61**, 57–66.

COLLIER, G., & SISKEL, M., JR. Performance as a joint function of amount of reinforcement and inter-reinforcement interval. *J. exp. Psychol.*, 1959, **57**, 115–120.

HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, 62, 243–254.

(Received August 1, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 385–394

PROBABILITY LEARNING IN 1000 TRIALS[1]

WARD EDWARDS

Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan

This paper reports a simple two-alternative noncontingent probability learning experiment with an unconventional feature: each *S* made 1000 consecutive predictions, making possible very detailed analysis of responses which occurred after learning was essentially completed.

Some abbreviations will be useful. The *S* predicts either L or R; after each prediction he observes either l or r. The probability that *S* will make prediction L on trial $t+1$ will be called p_t. The probability of l on any trial is a constant for any given *S*; it will be called π. The occurrence of a prediction will be called a *response*; the occurrence of a display of an event following a response will be called an *outcome*. An outcome follows each response; the nature of the outcome is independent of the nature of the response.

The interpretation of this experiment will focus on three issues:

1. *The probability matching hypothesis.* The probability matching hypothesis (PMH) asserts that the asymptotic probability of choice, p_∞ ($p_\infty = \lim_{t \to \infty} p_t$; it is assumed that this limit exists) equals π. It was originally proposed by Grant, Hake, and Hornseth (1951), is predicted by the Estes learning model (Estes, 1950, 1957; Estes & Burke, 1953; Estes & Straughan, 1954) and by the equal-alpha case of the Bush-Mosteller learning model (Bush & Mosteller, 1955), and has been supported by a number of experiments, though not by others.

2. *The extreme-asymptote generalization.* In 1956 I reported an experiment which argued against PMH and in favor of a theory about p_∞ which I call the extreme-asymptote generalization. That generalization asserts that p_∞ is more extreme than π, and as the absolute value of the difference between π and 0.5 increases the difference between p_∞ and π also increases until p_∞ becomes 1 or 0. As stated, this hypothesis makes only ordinal predictions; a way of making it yield ratio scale predictions (and also of applying it to situations in which amount of payoff is varied) is discussed in Edwards (1956) and applied later in this paper.

3. *Sequential dependencies, the gambler's fallacy, and path independence.* Stochastic learning theories often assume that the effects of events prior to a given trial are summarized in a set of probabilities for the responses available on that trial; this assumption is known as the path independence assumption (for a better definition, see Bush & Mosteller, 1955, p. 17). Contradictory to this is the common observation that if a flipped coin comes up heads eight or nine times in a row, *S* is likely to decide that tails is "due" and so predict or bet on it on the next toss. This and similar sequential effects have been called the gambler's fallacy; they have been demonstrated experimentally by many *E*s. Other hypotheses about sequential effects in probability learning have also

---

[1] This research was supported in part by the Operator Laboratory, Air Force Personnel and Training Research Center, and in part by Project MICHIGAN under Department of the Army Contract DA-36-039-SC-78801, administered by the United States Army Signal Corps. Reproduction for any purpose of the United States Government is permitted. I am grateful to L. W. Dildy and J. L. McDade for collecting and processing data, to W. H. Evans for writing computer programs, and to Carol J. Orwant and G. C. Morningstar for additional data processing.

been proposed, with varying degrees of empirical support.

Adequate study of each of these issues depends on long experiments; the reasons why will be examined in the discussion section. Both PMH and sequential dependencies are harder to examine at more extreme probabilities than at less extreme ones. So this experiment used only the probabilities 0.5, 0.6, and 0.7 and their complements.

METHOD

Apparatus.—Each *S* was given a tray containing 13 IBM mark sense multiple choice answer sheets. On top of the stack of sheets was a covering board with 80 pairs of holes in it, each hole filled by an ordinary cork. Each hole exposed two adjacent spaces where a mark could be made on the answer sheet. The mark sense sheets were prepared in advance by filling in one of the two mark spaces under the right-hand hole of each pair.

Subjects.—The *S*s were 120 basic airmen, trainees at Lackland Air Force Base. They were unselected except that no *S* who fell in Category 4 (the lowest category) of the Armed Forces Qualification Test, a paper-and-pencil test of general intelligence, was used. But the population of basic airmen includes relatively few college level men. The *S*s used in this experiment, therefore, are selected from a population which has almost no overlap with the college population from which *S*s have been selected for other probability learning experiments, except those by Neimark and Shuford (1959) and Nicks (1959), who also used basic airmen.

Instructions.—Each *S* was told to lift the upper left hand cork, and to make a mark in either the left or the right space on the sheet underneath it. A mark in, for instance, the left space was a prediction that the left space under the other cork of the pair would turn out to be filled in. After making the mark, he lifted the other member of the pair of corks, and saw whether his prediction had been correct or incorrect. After this, he replaced both corks, lifted the cork immediately beneath the first one he had lifted, and made his next prediction. When he finished 80 predictions, he removed the covering board, put the finished answer sheet underneath the stack, replaced the covering board, and continued making predictions. All *S*s were instructed: "Your purpose is to get as many predictions correct as possible. You will not be able to get all of them correct at any time during the test. There is no pattern or system you can use which would make it possible to get all of your answers correct. But you will find that you can improve your performance in the test if you pay attention and think what you are doing."

Experimental design.—There were 12 groups of 10 men each; each *S* made 1000 binary predictions in one unbroken session, usually lasting about 3 hr. The *S*s came in groups of 12; each *S* was arbitrarily assigned to one of the experimental groups. Twelve *S*s and *E* sat at a long conference table; *E* monitored continuously to make sure that all *S*s followed instructions and kept at the task. No effort was made or needed to pace *S*s. Each *S* present at a given time was a member of a different experimental group from all others then present, so no *S* could profit from looking at another *S*'s predictions.

Three basic probabilities were used: 0.5, 0.6, and 0.7. These numbers are the probabilities that a prediction of left will be correct. Sequences of 1000 trials embodying these probabilities were prepared in two different ways, which this paper will call *constrained* and *random*. All constrained sequences were prepared as follows. First, the expected number of occurrences of runs of length 1, 2, . . ., *n* for each of the two alternatives was calculated, up to a value of *n* for which that expected number is less than 0.5. All numbers were rounded off to integers. The runs of l were put in one box, the runs of r were put in another, and runs were drawn at random from the two boxes alternately until both were empty. This procedure makes not only run lengths but also conditional probabilities (based on sequences which are short compared with *n*) come out at their expected values. The random sequences were simply chosen from a table of random numbers in accordance with their probabilities, with no constraints at all.

Three probabilities and two ways of preparing sequences require six sequences. Six more sequences, each a mirror image of one of the six original sequences, were also used. The mirror image sequences were prepared by substituting an l for each r and an r for each l. One of these 12 sequences was administered to each of the groups; all *S*s in a group got the same sequence.

RESULTS

Asymptotic probabilities.—Figure 1 shows mean relative frequencies of choice by blocks of trials. Each data point represents 40 binary choices by each of 10 *S*s, or 400 binary choices in all. In each of the eight groups for which the probability of reward is not 0.5 and so for which PMH and the extreme-asymptote generalization make different predictions, the results support the extreme-asymptote generalization. Inspection of the 50–50 groups suggests that there is a bias in favor of the R response (which is surprising, since for a right-handed *S* the L response is a trifle easier to make), but the bias is not large enough to affect the finding.

Inspection of Fig. 1 indicates that *S*s tended to follow local changes in the probability of reward. A local increase in frequency of 1 events produces a local increase in frequency of L predictions, and similarly for decreases. This effect is superimposed on the slower and larger changes

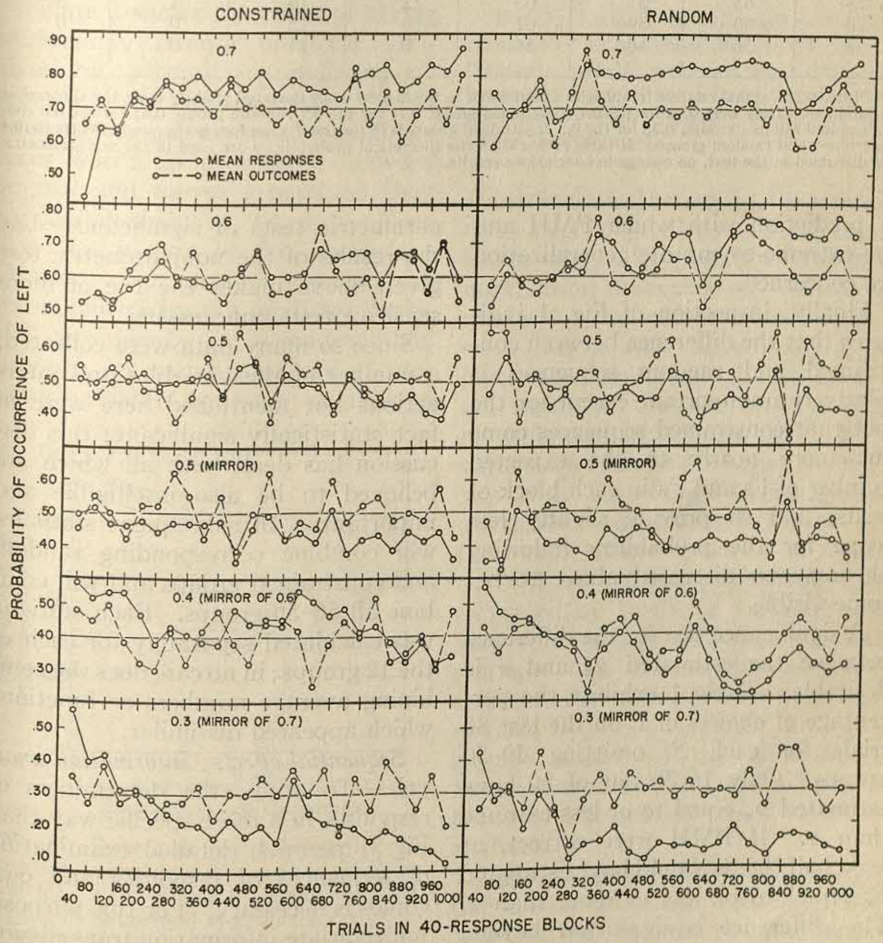


FIG. 1. Probability of left response in 40-trial blocks. (The number at the top of each rectangle is the π for that group, as is also the thin horizontal line within each rectangle. Each data point connected with solid lines is the relative frequency of prediction of left on a given block of trials; each point is based on 400 binary choices. Each point connected with dashed lines is the relative frequency with which the left event actually occurred in that block of 40 trials.)

TABLE 1

PERCENTAGE OF PREDICTIONS OF LEFT ON LAST 80 TRIALS FOR EACH S NOT IN A 50-50 GROUP

$\pi = 0.7$		$\pi = 0.6$		$\pi = 0.4$ (Mirrors of 0.6 Groups)		$\pi = 0.3$ (Mirrors of 0.7 Groups)	
Constrained	Random	Constrained	Random	Constrained	Random	Constrained	Random
100%	100%	76%	91%	48%	49%	26%	26%
97	93	74	90	47	43	20	20
96	88	70	85	46	40	17	13
95	88	69	81	43	35	12	13
91	87	66	77	43	31	11	13
85	85	64	74	31	26	8	13
80	80	64	71	29	21	4	11
75	65	61	63	22	16	4	8
70	60	59	61	20	15	0	0
58	56	46	56	0	11	0	0

Note.—The actual relative frequencies of outcomes in the last 80 trials deviated slightly from the theoretical probabilities. They were 0.73 for the 0.7 constrained group, 0.74 for the 0.7 random group, 0.61 for the 0.6 constrained and random groups, 0.28 for the 0.3 constrained group, 0.26 for the 0.3 random group, and 0.39 for the 0.4 constrained and random groups. If these rather than the theoretical probabilities are used in the nonparametric test discussed in the text, no change in conclusions results.

in prediction with which PMH and the extreme-asymptote generalization are concerned.

Finally, inspection of Fig. 1 indicates that the difference between constrained and random sequences is relatively unimportant except for the fact that constrained sequences come out more nearly to the expected number of l's and r's in each block of trials, and so provide slightly less scope for the probability following phenomenon discussed above to become visible.

A significance test for the difference between the estimated p_∞ and π is desirable. Table 1 exhibits the percentage of choices of L on the last 80 trials for each S, omitting 50–50 groups. Only 16 Ss out of 80 have estimated p_∞ equal to or less extreme than π. If PMH were correct, at least half the Ss should have estimated p_∞ equal to or less extreme than π. The difference is significant beyond the .0001 level. Table 1 also makes it clear that the distribution of estimated p_∞ is not bimodal; indeed, it looks relatively normal. That fact permits the use of more sensitive parametric tests of significance—but the results of the nonparametric test given above makes the use of more sensitive tests unnecessary.

Since so many data were collected, a number of the variables and interactions not mentioned here were in fact statistically significant; this discussion has dealt with all which are believed to be also intelligible and important. All subsequent statistics will combine corresponding random and constrained groups and will combine all 50–50 groups. Each statistic was calculated separately for each of the 12 groups; in no case does the combining average numbers or functions which appeared dissimilar.

Sequential effects: *Information analysis.*—To study the determiners of responses in a more specific way than Fig. 1 permits, detailed examination of sequences of responses and outcomes is necessary. For this purpose, multivariate information transmission analysis (Garner & McGill, 1956; McGill, 1954) is exceptionally convenient. The model underlying the use of this statistic assumes stable conditional probabilities; the analysis avoided basing calculations on changing overall probabilities by using only the last 480 trials. Special attention to the nonorthogonality of predictor variables and to the choice of proper degrees of freedom for the Miller-Madow (1954) bias correction and significance test was necessary; for a discussion of these issues and related ones concerning the application of information statistics to sequences of responses, see Edwards (1954; in press).

Figure 2 shows the effect of taking increasingly remote predictor variables into account in predicting responses in the last 480 trials. (In all information calculations, no differences worth noting existed between original and mirror groups, so they are combined in Fig. 2 and 3.) Note that the y axis is the percentage of information in the responses *not* accounted for by the predictor variables considered. It is evident that although increasing numbers of predictor variables improve predictions (a mathematical necessity), the asymptotic level of predictive effectiveness leaves about 75% of the response information unexplained. If these numbers were variances, this would seem like a very large amount of unexplained variance. But they are not variances; they are ratios of bits of information. Users of multivariate information transmission analysis always report large percentages of unexplained response information; in fact, experiments in which as much as 25% of response information is explained by predictor variables are very rare (except in psychophysical scaling). No formal discussion of this common finding is known, but an obvious hypothesis is that the logarithmic nature of the information measure accounts for this difference between information and variance analyses.

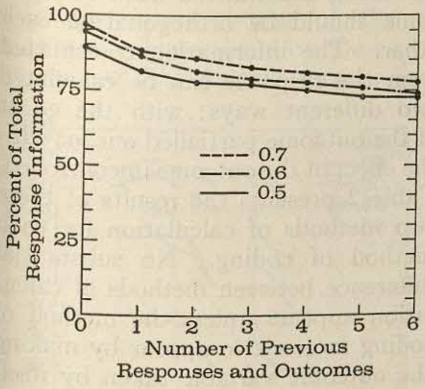


FIG. 2. Percentage of total response information unexplained by various predictor variables. (The x axis is cumulative. At Step 0, only *S*s are used as predictor variables. At Step 1, *S*s and the immediately preceding outcome are used. At Step 2, the variables already listed and also the immediately preceding response are used. At Step 3, the variables already listed and also the second preceding outcome are used. And so on. Only the last 480 trials were used.)

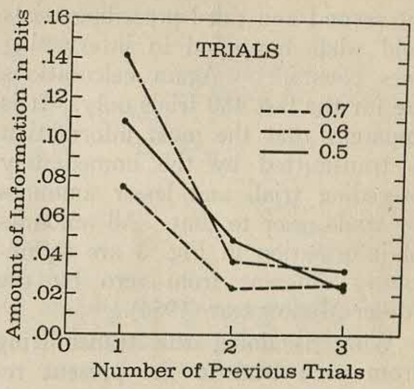


FIG. 3. Amount of information in bits transmitted by preceding trials to the present response. (The *S*s and trials which intervene between the predictor trial and the predicted response are held constant. Only the last 480 trials were used.)

Figure 2 shows how much prediction can be done, but does not show how to do it. In order to get a better idea about that, consider Fig. 3. It shows the amount of information (in bits, not a ratio) transmitted to the present response by the preceding three trials (calculations for the second and third preceding trials hold what happened in intervening ones constant). Again calculations are for the last 480 trials only. It is apparent that the most information is transmitted by the immediately preceding trial, and lesser amounts by trials prior to that. All amounts of information in Fig. 3 are significantly different from zero by the Miller-Madow test (1954).

What is doing the transmitting from each trial to the present response? It could be responses, outcomes, interactions between them, or any combination of these three factors. Unfortunately, the interactions between responses and outcomes are not directly interpretable because of the nonorthogonality of the predictor variables. Figure 2 is based on a definition of outcomes as being l or r; call this *noncontingent* coding. It would also be possible to define outcomes as + or − (meaning in agreement or disagreement with the preceding prediction); call this *contingent* coding. Further analysis of the data using noncontingent coding shows that almost all information transmitted by a trial is transmitted by its outcome; the amount of information transmitted by responses is trivial (though significant; because of the large numbers of responses involved, just about all differences which are observable at all are significant in this experiment). The implication, a sensible one, is that *S*s pay little or no attention to their own previous responses and instead concentrate on the previous set of outcomes in determining their present response.

Of course, similar analysis applied to contingently coded data shows that almost all information transmitted by a trial is transmitted by its response; this is an inevitable consequence of the fact that a + or − is meaningless as a predictor variable unless the preceding response which defines it is also considered. So two different methods of coding the data lead to two different interpretations of the results. A decision between these interpretations would require examination of the interactions, and nonorthogonality rules out the obvious ways of doing so. But a stab at it is available. If only the trial immediately preceding a response is considered, then the information transmitted from the response and information transmitted from the outcome should be orthogonal to each other. The information transmitted from the response can be calculated two different ways: with the effect of the outcome partialled out, or with the effect of the outcome uncontrolled. Table 2 presents the results of these two methods of calculation for each method of coding. No substantial difference between methods of calculation appears unless the method of coding forces it to appear by making the outcome variable taken by itself meaningless. For that reason, this paper used the noncontingent method of coding, and will accept the conclusion that *S*s are much more concerned with previous outcomes than with their own previous responses. Conclusive resolution of the dilemma, however, would require a three-alternative experiment, in which case

TABLE 2

INFORMATION IN BITS TRANSMITTED FROM PREVIOUS RESPONSE TO PRESENT RESPONSE BY THREE METHODS OF CALCULATION

Analysis	π		
	0.5	0.6	0.7
Intervening outcome ignored	.024	.040	.016
Intervening outcome held constant, noncontingent coding	.031	.045	.020
Intervening outcome held constant, contingent coding	.104	.131	.058

contingent and noncontingent coding would not in general lead to the same amounts of information transmission.

Sequential effects: *Run analyses.*—The information statistics presented above examine sequential effects in a manner which assumes that the extent of sequential dependency is independent of the particular sequence considered. Clearly that assumption can be only a first approximation. The literature suggests that one kind of past history is especially likely to lead to sequential effects: homogeneous runs of previous outcomes. Rather than examine such runs by information methods, it is easier to examine conditional probabilities based on them directly. Figure 4, again based on the last 480 trials only, shows the conditional probability (multiplied by 100) that L will be predicted given each possible preceding homogeneous outcome run of length eight or less. The data do not permit these probabilities to be estimated for longer runs with acceptable accuracy. An example may make the interpretation of the x axis easier. The value 4 on the right run side of the x axis means, for example, that the points plotted above it are conditional probabilities of predicting L given that the last five outcomes preceding the prediction were lrrrr. (Note that one actually knows five preceding outcomes, not four, since the outcome preceding a homogeneous outcome run of r must necessarily be l, and vice versa.) Figure 4 justifies the conclusion that outcome runs of length up to four certainly influence responses, and so indicates that for at least some past histories the extent of sequential dependencies is longer into the past than the information analysis taken alone would suggest. But the nature of the dependencies is that the longer an outcome run gets, the more likely S is to predict that outcome. What happened to the gambler's fallacy?

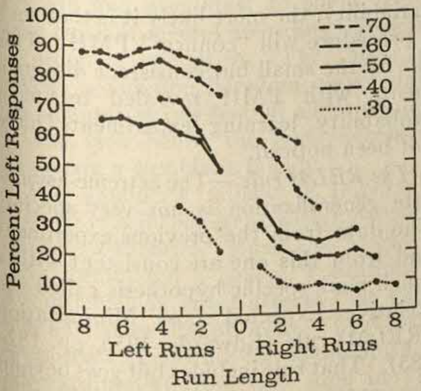


FIG. 4. Percentage of left responses following homogeneous outcome runs. (The x axis indicates the number of left or right outcomes included in the run for further explanation, see text. Only the last 480 trials were used.)

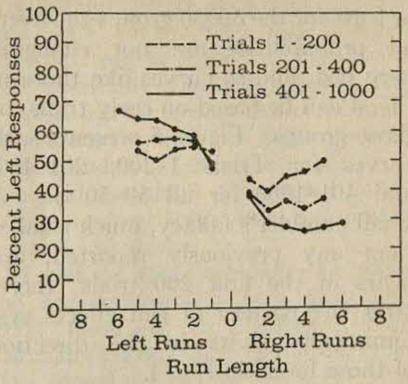


FIG. 5. Percentage of left responses following homogeneous outcome runs for 50–50 group Ss only. (The axes have the same meaning as those in Fig. 4.)

Most experiments which have found gambler's fallacies used fewer trials than this one. Perhaps the gambler's fallacy is a phenomenon of early trials and vanishes later. If so, strictly speaking no run curves like those in Fig. 4 are appropriate to use in studying it during early trials, while response probabilities are changing rapidly. But it is reasonable to assume as a first approximation that at least for the 50–50 groups the overall probabilities are not changing very fast, and so curves like those in Fig. 4 can be based on early trials for those groups. Figure 5 presents such curves for Trials 1–200, 201–400, and 401–1000 for all 50–50 *S*s. A small gambler's fallacy, much smaller than any previously reported, appears in the first 200 trials; thereafter the pattern of run effects systematically shifts in the direction of those found in Fig. 4.

DISCUSSION

Probability matching.—In 1956 I reviewed all experiments relevant to a narrow definition of PMH published up to that time (Edwards, 1956, pp. 184–185). Only experimental groups in which the two outcomes were mutually exclusive and exhaustive, in which successive outcomes were independent, in which π was not 0, 0.5, or 1, and in which *S* had had no previous experimental experience with a different value of π were considered. Of 11 groups meeting these conditions, only 1 had an estimated p_∞ which was equal to or less extreme than π. In the other 10 groups, p_∞ was always more extreme than π. The differences were small, but they were all in the same direction.

Of experiments containing relevant groups published since then, those by Gardner (1957), Cotton and Rechtschaffen (1958), and Nicks (1959) are inconsistent with PMH; those by Neimark (1956), Engler (1958), Neimark and Shuford (1959), and Rubinstein (1959) support PMH. No probability learning experiments (as here narrowly defined) reviewed in 1956 or published since then used more than 300 trials at a fixed probability except those by Gardner (1957), Cotton and Rechtschaffen (1958), and Nicks (1959), all three of which are inconsistent with PMH. Figure 1 indicates that in this experiment probabilities of choice were still becoming more extreme at Trial 300 and beyond. Longer experiments at fixed π values might perhaps have produced fewer acceptances of PMH.

Why did PMH, at best dubiously supported by experimental data, achieve such widespread acceptance as a well-established truth? Three reasons seem plausible. First, it is a good first approximation to the truth. It is more nearly correct than the assertion that $p_\infty = 0.5$ for any value of π, or that $p_\infty = 1$ whenever π is greater than 0.5. Furthermore, it is predicted by some (not all) stochastic learning models, which themselves are good first approximations to the truth. Secondly, few experiments have run enough trials to obtain a reasonable estimate of p_∞. Inclusion of trials on which p_t is still changing substantially as a function of *t* in estimates of p_∞ will, of course, produce estimates of p_∞ which are less extreme than they should be, and so come closer to supporting PMH than they should. (The use of cumulative relative frequency as an estimator of p_∞, as in Estes [1957], will of course bias the estimates in favor of PMH still more.) Finally, the custom of obtaining an estimate of p_∞ and testing the null hypothesis that that estimate is not significantly different from π is widespread in the probability learning literature (and was done in this paper). Such a procedure constitutes attempting to prove a null hypothesis; the smaller the amount of data or the greater its variability, the more likely it is that such a procedure will "confirm" PMH. This is why the small but consistent disagreements with PMH revealed by most probability learning experiments have not been noticed.

The RELM rule.—The extreme-asymptote generalization is not very specific. The data from the previous experiment and from this one are consistent with a much more specific hypothesis called the Relative Expected Loss Minimization (RELM) rule (Edwards, 1956, pp. 182–185). That rule includes but goes beyond the extreme-asymptote generalization, and is applicable to a wide variety of experiments. For this kind of experiment, the linear form of that rule predicts that $p_\infty = 0.5 + K(4\pi - 2)$, where

K is a fitted constant greater than 0.25. The size of K presumably varies with motivational and other characteristics of the experimental design. A least squares fit shows that for the data obtained in this experiment $K = 0.395$.

Sequential effects.—The surprise in this experiment is the weakness of the gambler's fallacy found, and its disappearance in later trials. Nicks (1959), Anderson (1960), and Anderson and Whalen (1960) found much larger gambler's fallacies in appropriate groups; in fact, Anderson found gambler's fallacy effects even when his sequences were designed so that the probability of an outcome repetition was higher than it would have been had successive outcomes been independent. (Jarvik [1951] also found large gambler's fallacies, but his experiment was so designed that they were not at all fallacious.) But this experiment does not stand alone; Feldman (1959b) found no gambler's fallacy at all in his 200-trial experiment.

No real explanation of this divergence in presumably similar experiments is apparent. It is possible, however, that the relative inconvenience of the responses in this experiment served to increase the monotony of what was in any case an exceedingly monotonous task. The gambler's fallacy is in a sense a highly intellectual response. The S must have some idea of what probabilities are and also must to some degree keep track of several preceding outcomes in order to exhibit it. For this non-college population boredom may reduce the amount of intellectual effort applied to the task below the level necessary to sustain a gambler's fallacy.

The gambler's fallacy is important because it is inconsistent with most reinforcement theories. Bush and Morlock[2] have formulated a general conditioning axiom which in effect asserts that gambler's fallacies cannot occur. They have proposed a procedure for examining run effects different from that used in Fig. 4 and 5; they examine only the responses and outcomes included in outcome runs of a specified length (or longer). These data were analysed by their method for Run Lengths 5 and 7. The results were essentially similar to those in Fig. 4 and 5, but the greatly decreased number of observations per point resulted in a considerable decrease in stability. The evidence about the general conditioning axiom from this experiment remains ambiguous.

Hypothesis-testing behavior.—Goodnow and her collaborators (e.g., Goodnow, 1955; Goodnow & Postman, 1955), Feldman (1959a), and I (Edwards, 1956) have argued that people base predictions in probability learning on local hypotheses about sequential dependencies. This idea is very attractive; the sequential effects examined in this paper make it more so. Unfortunately, too many hypotheses (most necessarily incorrect) are possible, and they change too fast and too irregularly, to make this an easy idea to use. Feldman, working with verbal statements as well as predictions, has found it necessary to construct one hypothesis per S. This is the end point of any attempt to give a detailed, explicit account of probability learning from a hypothesis-testing point of view. We need higher order models, so that each specific set of hypotheses can be included within some more general classificatory or explanatory scheme. No such models are available at present.

SUMMARY

A probability learning experiment is reported in which each of 120 Ss made a sequence of 1,000 predictions about which of two mutually exclusive events will occur. After each prediction, one of the two events occurs; the probability of occurrence of each event is constant (0.5, 0.6, 0.7 and their mirror images). Sequences were randomized in two different ways. For all relevant groups, the asymptotic probability of prediction was more extreme than the probability of occurrence of the event predicted; probability matching did not occur. The Ss responded to small increases or decreases in the relative frequency of an event in a block of trials by similar small increases or decreases in their predictions of that event in that block; this phenomenon was named probability following.

---

[2] Bush and Morlock, personal communication.

Examination of sequential dependencies by means of information measures indicates that about 25% of response information can be accounted for by the identity of *S*s and the results of the last three trials. The *S*s apparently pay most attention to previous outcomes, and much less attention to previous responses. Most of the predicting is done by the immediately preceding trial; trials further back contribute only small amounts of additional transmitted information.

Analyses of homogeneous outcome runs on later trials show that the longer the run of occurrences of an event, the more likely *S* is to predict that event. For early trials, however, *S*s show a slight tendency to predict the event less often as its run length increases; this is the gambler's fallacy.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

ANDERSON, N. H., & WHALEN, R. E. Likelihood judgments and sequential effects in a two-choice probability learning situation. *J. exp. Psychol.*, 1960, **60**, 111–120.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

COTTON, J. W., & RECHTSCHAFFEN, A. Replication report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

EDWARDS, W. Methods for computing uncertainties. *Amer. J. Psychol.*, 1954, **67**, 164–170.

EDWARDS, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

EDWARDS, W. Measure of information based on overlapping counts for studying sequential dependencies. *Amer. J. Psychol.*, in press.

ENGLER, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

FELDMAN, J. An analysis of predictive behavior in a two-choice situation. Unpublished doctoral dissertation, Carnegie Institute of Technology, 1959. (a)

FELDMAN, J. On the negative recency hypothesis in the prediction of a series of binary symbols. *Amer. J. Psychol.*, 1959, **72**, 597–599. (b)

GARDNER, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

GARNER, W. R., & MCGILL, W. J. The relation between information and variance analyses. *Psychometrika*, 1956, **21**, 219–228.

GOODNOW, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

GOODNOW, J. J., & POSTMAN, L. Learning in a two-choice probability situation with a problem-solving setting. *J. exp. Psychol.*, 1955, **49**, 16–22.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1951, **42**, 1–5.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

MCGILL, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97–116.

MILLER, G. A., & MADOW, W. G. On the maximum likelihood estimate of the Shannon-Wiener measure of information. *USAF CRC tech. Rep.*, 1954, No. RE-54-75.

NEIMARK, E. D. Effect of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

NEIMARK, E. D., & SHUFORD, E. Comparison of predictions and estimates in a probability learning situation. *J. exp. Psychol.*, 1959, **57**, 294–298.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

RUBINSTEIN, I. Some factors in probability matching. *J. exp. Psychol.*, 1959, **57**, 413–416.

(Received August 4, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 395–402

PRIMARY AND SECONDARY STIMULUS GENERALIZATION AS RELATED TO INTELLIGENCE LEVEL[1]

IRV BIALER[2]

George Peabody College

According to S-R theory (Miller, 1948), learning to respond with distinctive labels (cue-producing responses) to two similar stimulus situations should tend to increase the difference between them and thus facilitate subsequent discriminations. Increased differentiation based on this mechanism has been called "acquired distinctiveness of cues" (Miller, 1948, p. 174). On the other hand, if the individual learns to attach the same verbal response to two distinctive stimulus situations, the resulting response-produced cues (e.g., "same name") give these situations a certain amount of learned equivalence. The transfer of an instrumental response from one such secondarily equivalent situation to the other is said to be mediated by the "acquired equivalence of cues" (Miller, 1948, p. 174)—or secondary generalization.

Again, current S-R formulations (Dollard & Miller, 1950; Miller, 1948) hold that it is primarily the verbal cue-producing responses which mediate man's higher mental processes; and the absence or removal of these learned responses (such as occurs in repression) results in relatively unintelligent behavior.

The point is also made by these same theorists that the unavailability of learned verbal cue-producing responses results not only in less intelligent behavior, but also in more primary stimulus generalization (response to superficial sensory similarity of stimuli which have been given otherwise distinctive characteristics in the form of different labels) where discrimination is required, and in less secondary stimulus generalization where the utilization of learned equivalence is required.

The present study is designed to investigate the postulated relationship between verbal cue-producing responses, intelligent behavior, and primary vs. secondary stimulus generalization.

It was expected that when groups of *S*s differing in intellectual ability are compared in a transfer situation under combined conditions of acquired distinctiveness and acquired equivalence of cues, both primary and secondary stimulus generalization may occur over the entire intellectual range studied. However, as one goes up the range, relatively more secondary generalization responses should be found; and as one goes down the intellectual scale, relatively more responses based on primary stimulus generalization (consequent to a relative inability to retain the basis for acquired distinctiveness and equivalence) should be obtained.

---

[1] This study was conducted while the author was on internship at Southbury Training School, Southbury, Connecticut, during 1959–60. The study was partially supported by a grant (M-647) from the National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland; and grateful acknowledgment is made to Neal E. Miller, Yale University, for his aid and encouragement as research supervisor during the internship period, and for critical reading of drafts of the paper. The writer particularly wishes to express his appreciation to C. Edward Stull, Director of Psychological Service, and to the entire staff of the psychology department of Southbury Training School for their help and advice during the conduct of this study. Shirley Steele was especially helpful in screening the mongoloid *S*s; and David Barron contributed greatly in preparing the geometric designs utilized as stimuli.

[2] Now at Tennessee Clover Bottom Home, Donelson, Tennessee.

The specific hypothesis to be tested was: There is a significant relationship between intelligence level and transfer task response, such that higher IQ groups tend to give relatively more responses based on secondary stimulus generalization than do comparatively lower IQ groups.

METHOD

Subjects

The *S* sample consisted of 48 mentally retarded and normal boys. Twenty *S*s were of normal intelligence, ranging in IQ from 91 to 109 (mean IQ = 99.8), as measured by the Otis Quick-Scoring Mental Ability Test. These *S*s came from Grades 9 through 12 at the Southbury High School,[3] and they ranged in CA from 14-2 to 17-10 (mean CA = 16-4). The retarded *S*s (20 high grade familials and 8 middle grade mongoloids) were residents at Southbury Training School, a state institution for mental defectives. The high grade group ranged in IQ from 51 to 75 (mean IQ = 62.6) and in CA from 14-1 to 17-9 (mean CA = 16-1). With the exception of 3 *S*s, who had been tested with the Wechsler Adult Intelligence Scale, the Wechsler-Bellevue Intelligence Scale, and the revised Stanford-Binet, Form L, respectively, IQs for the high grade sample were derived from the Wechsler Intelligence Scale for Children. The middle grade *S*s had an IQ range of 26 to 37, with a mean IQ of 31.8 (revised Stanford-Binet, Form L), and they ranged in CA from 14-1 to 29-1 (mean CA = 21-8). In the case of the latter group, it was necessary to go beyond the CA range used in the other groups in order to find mongoloids whose speech and comprehension ability were adequate for the required experimental tasks.

In selecting all *S*s for inclusion in the study, an effort was made to screen out children with debilitating physical and emotional defects.

Materials and Procedure

Two different sets of stimuli were employed. One set consisted of four geometric designs (form stimuli), individually drawn on top of otherwise identical square box covers (Boxes 1, 2, 3, and 4). The designs, shown in Fig. 1, were modified from those used by Cantor and Hottel (1957). The other stimulus set (color stimuli) consisted of four box covers (Boxes 1A, 2A, 3A, and 4A) which were covered by colored construction paper. Boxes 1A and 2A were light and dark blue, respectively; Boxes 3A and 4A were light and dark red, respectively. Each of the eight stimulus boxes measured 3 × 3 × ¾ in.

The normal and high grade retarded *S*s received both sets of stimuli (Form and Color conditions) in succession. To control for the possible effect of order of stimulus presentation on performance in the pretraining and transfer situations, a random half of each of these groups was presented with the tasks involving Form first, the other half receiving the Color tasks first.

However, on the basis of the performance of the above *S* groups with the Color stimuli, it was evident that the mongoloids would find coping with these stimuli extremely difficult, if not impossible. Therefore, the

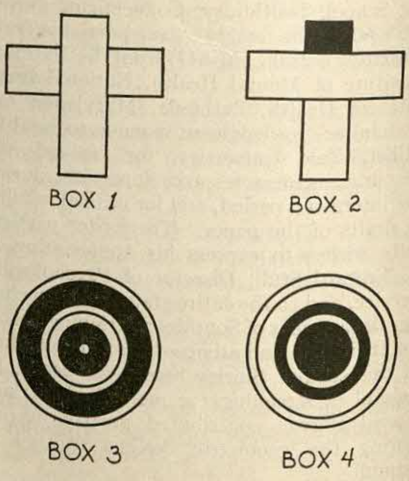


FIG. 1. Geometric designs used as form stimuli.

---

[3] The author is grateful to Thomas J. Pepe, Superintendent of Schools, Southbury, Connecticut, and to James H. Heller, Director of Guidance, Southbury High School, for their generous assistance and full cooperation in providing these *S*s.

middle grade retardates were exposed to the Form conditions only. For all *S*s, the following procedure was introduced as a "name-learning game."

Form conditions.—The pretraining situation consisted of two major phases: (*a*) four name-learning tasks (Cond. IA, IB, II, and III); and (*b*) a candy-finding task (Cond. IV).

In the name-learning phase, each *S* was first required to learn to associate the name "bif" to Box 1 and the name "mot" to Box 2 (Cond. IA). The *S* then learned to call Box 3 bif and Box 4 mot (Cond. IB). Under each condition, *S* was shown the pair of boxes together and told the name that went with each of the boxes. The stimuli were then presented singly, in a predetermined random sequence. The *S*'s errors were corrected by *E*, and training was continued until *S* responded correctly to five out of five consecutive presentations. When this criterion was reached, it was assumed that *S* had learned to distinguish Box 1 from Box 2 and Box 3 from Box 4 according to the principle of acquired distinctiveness of cues. It was also assumed that, since *S* had apparently learned the same name (bif for Boxes 1 and 3 and mot for Boxes 2 and 4) for quite different stimuli, each given pair of cues had now acquired some degree of equivalence for *S*.

However, in order to make the foregoing assumptions more tenable, *S* was next presented with a pack of 20 cards, made up of five replicas of each of the four box tops (Cond. II). The *S* was told that the cards had the same designs and the same names as the boxes; that he was going to be shown the cards one at a time; and that each time, he was to tell *E* the name of the card. The cards were shown single, in a predetermined nonalternating sequence. The *S*'s errors were again corrected by *E* until *S* reached a criterion of 9 correct responses on any 10 consecutive presentations.

At this point, it was assumed that the dual conditions of acquired distinctiveness and equivalence of cues had been well established. Nevertheless, since the subsequent candy-finding and transfer tasks involved the boxes, rather than the cards, the preceding procedure was repeated with the four boxes (Cond. III). The *S* was instructed that he was now going to be shown the four boxes again and that he was to tell *E* the name of each box as it was presented. The boxes were shown, one at a time, in a predetermined random sequence, and *S*'s errors were corrected until he spontaneously named the boxes correctly on eight consecutive trials.

Under Cond. IV, in pretraining, *S* was required to learn to associate a manipulative response (reaching for and picking up one of a pair of boxes to find a candy reward) with a verbal cue-producing response, involving the name of the particular box covering the candy. The procedure for this condition was a modification of that used by Birge (1941).

Multicolored M & M candies were used as reinforcement, and a random half of each group was trained to find the reward under Box 1 (bif), the other half learning to find the candy under Box 4 (mot). In presenting the stimulus pairs to *S*, the left-right positions of the boxes were randomized for each *S* from trial to trial. All *S*s received the same L-R sequence.

The *S* was shown Boxes 1 and 4 (with relative L-R positions opposite to that in which they were subsequently to appear on the first learning trial) on a turntable, and *E* said:

> This time, you are going to get some chances to find candy under one of these boxes. Each time, the candy will be under the bif (or mot). When I show you the boxes, I want you to say, "The candy is under the bif (or mot)," and then, pick up the bif (or mot), and take the candy. You may eat the candy now, or save it for later in this envelope (*S* was given a small envelope). Be sure to get it right. First, you say, "The candy is under the bif (or mot)," and then, you pick up the bif (or mot) and take the candy. Ready? O.K.

For the first two learning trials, *S* was allowed to see the candy (a single M & M) being placed under the appropriate box. The turntable was then revolved, and, while it was still turning, it was hidden from *S*'s view behind a cardboard screen. After the screen was in place, the turntable was stopped; the boxes were moved to their predetermined L-R positions; and *E* asked: "Where is the candy?" On these trials, as on all subsequent ones, *S* was required to change incorrect verbal and/or manipulative responses before being allowed to take the candy. Thus, each trial ended with *S* being finally rewarded for a series of correct responses.

In an effort to insure that *S*'s manipulative response would become associated primarily to the verbal response, the procedure was modified beginning with the third candy-finding trial. Now the screen was placed between *S* and the revolving turntable before the candy was placed by *E*. In this way, *S* had no further opportunity to utilize possible visual cues provided by *E*'s manipulation of the appropriate stimulus object; and trials were continued until *S* responded correctly without the candy-placement cue from *E* on five consecutive presentations of the paired stimuli. Thus, the lowest score any given *S* could achieve in Cond. IV was seven trials to criterion.

When *S* had attained criterion on the preceding task, the transfer situation was introduced without a break in the procedure and without giving *S* any warning that changes were being made in the stimulus conditions. Having placed the screen, *E* removed Boxes 1 and 4 from the turntable (as had been done prior to all previous trials). The original stimuli were then replaced by Boxes 2 and 3, relative L-R positions of this pair having been randomized from *S* to *S*, and the candy was again placed under the box with the appropriate name. With the removal of the screen, *S* was again required to choose the box with the candy by giving the necessary verbal and manipulative responses.

Each *S* was given only one transfer trial, and, depending upon whether he reached for and picked up the box having the same name as, or the box bearing the geometric figure similar to, that under which the candy was found in pretraining, *S* was scored as having chosen "Name" or "Form," respectively. This score was taken as indicating whether the manipulative response had been mediated by secondary stimulus generalization, on the one hand, or primary stimulus generalization, on the other; and the number of *S*s achieving the respective scores formed the basis for testing the experimental hypothesis.

Color conditions.—The pretraining and transfer procedures (along with the appropriate controls) involving the color stimuli paralleled the above methodology exactly. Boxes 1A, 2A, 3A, and 4A appeared in situations analogous to those where Boxes 1, 2, 3, and 4 were used. However, in addition to differing from the form stimuli in physical appearance, the color stimuli carried a different set of verbal labels.

Briefly, during pretraining, *S* first learned to associate the name "dap" to Boxes 1A (light blue) and 3A (dark red), and the name "guz" to Boxes 2A (dark blue) and 4A (light red). He was then shown Boxes 1A and 4A and learned to pick up one of these for candy, at the same time reciting the cue-producing response: "The candy is under the dap (or guz)." On the transfer task, *S* was scored as having made a "Name" or "Color" response, and the statistical analysis was based on the number of *S*s scoring in one category or the other.

In an effort to minimize communication from those *S*s who had already been seen to those yet to come, *S* was not released immediately after the main procedure was completed. Instead, *S* was shown four boxes, with the picture of a different animal (bison, tern, sheep, and owl) on each box, and he was told: "Now this is the really important part. Where is the candy? Pick up the box with the candy." Each box covered five pieces of candy. Thus, *S* was amply rewarded for whichever choice he made. It was hoped that this addendum to the procedure would serve as an effective "smokescreen" to *S*'s memory of what had come before. Nevertheless, before being allowed to leave, *S* was assured that *E* would appreciate it if the other boys were not told exactly what had taken place during the "game."

RESULTS

The statistical treatment was chiefly concerned with a chi square analysis of the data on the transfer task. Nevertheless, comparisons were also made between the various groups as to trials to criterion under both phases of the pretraining situation. Since the middle grade retardates had worked with the form stimuli only, statistical treatment of all the mongoloid group data consisted of comparisons of those data with responses made under the Form conditions by the other Form-first subgroups only.

In all the following analyses (pretraining and transfer tasks), the .05 level was used for determining statistical significance, and two-tailed tests were employed throughout.

Pretraining

Name-learning tasks.—Each *S*'s scores for Cond. IA, IB, II, and III were combined into a single trials-to-criterion score for the given *S*. The data for the normals and high grade retardates were analyzed by an analysis of variance design (Lindquist Type III, 1953) involving two variables (normals vs. retardates, Form vs. Color, Form-first vs. Color-first) in each of three respective dimensions (intelligence level, stimulus condition, stimulus order) with the following results: (*a*) there were no significant interactions; (*b*) the retardates took significantly more trials to criterion over-all than did the normals—indicating that name learning was, in general, more difficult for the retarded *S*s; (*c*) both groups took significantly more trials to criterion with Color than with Form; and (*d*) the effect of stimulus order was not significant.

Individual *t* test comparisons of the middle grade retarded group with the other Form-first subgroups indicated that the normals learned the names of the form stimuli in significantly fewer trials than did the mongoloids. However, the high grade and middle grade retardates did not differ significantly from each other in the mean number of trials needed to learn the names of the geometric designs.

Candy-finding task.—The trials to criterion were analysed for the data obtained in Cond. IV, with the following findings:

1. In comparing the normals and high grade retardates, it was found that: (*a*) There were no significant interactions. (*b*) There was no significant difference in mean number of trials to criterion under the different stimulus conditions. From this, it may be inferred that once the names had been learned to both sets of stimuli, learning to associate a manipulative response to those names was, in general, no more difficult for Color than for Form. (*c*)There was a significant over-all difference between the IQ groups—the retardates apparently having more difficulty than the normals in learning the appropriate associations. (*d*) The effect of stimulus order was significant. The Color-first group as a whole (combined Color-first normals and Color-first retardates) took more trials than did the Form-first group as a whole.

2. The *t* test comparisons between the middle grade group and the other Form-first subgroups revealed that the mongoloids apparently did not differ significantly from either of the other groups in the mean number of trials needed to learn the appropriate associations.

Transfer

Order of presentation.—The effect of stimulus order on *S*s' transfer-task responses was examined by a series of chi square comparisons of nonindependent proportions (McNemar, 1955). In this way, tests were made of the significance of overall changes from the first to the second set of responses (where both sets of stimulus conditions had been administered) in the various IQ groups and stimulus order subgroups. All comparisons involved a correction for continuity, according to the formula provided by McNemar (1955, p. 231) for correcting chi squares computed from fourfold tables in which the frequencies or proportions involved are based on the *same S*s. In all cases, the hypothesis was that the proportion of *S*s giving a "Name" response on the first transfer task was equal to the proportion giving that response on the second transfer task. Table 1 presents a summary of the results of the various tests of that hypothesis.

The nonsignificant chi squares in Table 1 indicate that, regardless of which stimulus condition was presented first, both IQ groups were consistent in their responses from the first transfer task to the second;

TABLE 1

Tests for Significance of Changes from First Set vs. Second Set of Responses within Given Groups and Subgroups

Group	N	Chi Square	P
Normals	20	<1.00	>.05
Retardates	20	<1.00	
Form-first Normals	10	<1.00	
Form-first Retardates	10	3.20	
Color-first Normals	10	<1.00	
Color-first Retardates	10	<1.00	

i.e., the tendency to give a "Name" response was not significantly affected by stimulus order.

The experimental hypothesis.—Since there was no significant order effect, the chi square comparisons of the normals vs. high grade retardates were conducted without taking stimulus order into account. In the case of these groups, the transfer task data were examined separately for Form and Color conditions, with the following findings: (*a*) Under the Form conditions, 20 normals and 13 retardates gave the "Name" response. None of the normals and 7 retardates chose "Form." ($\chi^2 = 6.23$, corrected for continuity, with $P = .02$.) (*b*) Under the Color conditions, 16 normals and 5 retardates made the "Name" response; while 4 normals and 15 retardates chose "Color." ($\chi^2 = 12.13$, $P = .001$.)

As with previous analyses, the responses of the middle grade retardates were compared only with those made under the Form conditions by the other Form-first subgroups. These comparisons revealed the following: (*a*) Of the 10 Form-first retardates, 6 *S*s gave the "Name" response; while 4 *S*s chose "Form." In the middle grade group, none of the *S*s chose "Name," all 8 mongoloids made the "Form" response. ($\chi^2 = 4.80$, corrected for continuity, $P = .05$.) (*b*) Among the Form-first normals, all 10 *S*s made the "Name" response, with none choosing "Form." Comparison of this group with the middle grade retardates yielded a χ^2 of 14.10 (corrected for continuity), with $P = .001$.

Discussion

The findings clearly indicate that the experimental hypothesis was confirmed. Thus, the results on the transfer tasks seem to validate the proposed theoretical relationship between intelligent behavior, verbal cue-producing responses, and primary vs. secondary stimulus generalization.

In general, the findings of the present study are consistent with those of other research (e.g., Birge, 1941; Cantor & Hottel, 1957; Golu, 1958).

Birge investigated the acquired equivalence of cues formulation with children of normal intelligence. She found that generalization of a manipulative response from one stimulus to another was facilitated when the children had previously learned to give the same names to the two very different stimuli. The study by Cantor and Hottel was concerned with the applicability of the acquired distinctiveness of cues formulation to the learning behavior of mental defectives. The obtained results indicated that teaching the retardate different names for similar stimuli aided the learning of a differential motor response to the stimuli. Golu, in a study summarized by Brackbill (1960), found that a differential verbal response, in addition to a motor response, decreased the number of errors in discrimination learning in mentally defective Russian children.

In addition to being consistent with the cited research, the present investigation goes further by pointing up the differential role played by verbal cue-producing responses in the higher mental processes—as well as the importance of the context within which these processes are operating.

The nature of the present stimulus conditions was such that the similar geometric figures were apparently more distinctive than were the similar colors. Nevertheless, the results indicate that the normal group showed relatively greater ability to utilize verbal cues for secondary generalization (i.e., demonstrated a relatively higher level of reasoning ability) under both stimulus conditions than did the high grade retardates. However, a comparison of the responses made by the latter group under Form and Color conditions, respectively, indicates that, for this group, the effectiveness of verbal cues in mediating secondary generalization was related to the relative distinctiveness of the form vs. the color stimuli. On the other hand, the comparison of the middle grade retardates with the other Form-first subgroups seems to emphasize that the mongoloids were comparatively even more limited to responding in terms of the primary, "given" aspects of the stimulus conditions—an apparent indication of very low-level reasoning ability.

It should be pointed out that although all the mongoloid *S*s gave the "Form" response, 3 of these *S*s had apparently been able to apply the correct name to the manipulated stimulus, even though they did not use this name to mediate transfer. For example, if they had been saying: "The candy is under the bif (cross)" in pretraining, when the boxes were switched for the transfer situation, these *S*s said: "The candy is under the mot (different cross)," and they picked up the mot. These *S*s picked up the physically more similar box in spite of the fact that they correctly gave it the different name (which had, for them, never been associated with finding the candy). Thus, for these *S*s, the "incorrect" manipulative response on the transfer task was made in spite of a correct verbal label.

However, all of the other *S*s, in all IQ groups, who transferred on the basis of physical similarity (primary stimulus generalization) not only reached for the "wrong" box, but also gave it the wrong name. These *S*s first gave the correct verbal response (e.g., "The candy is under the bif")—following which, they picked up the box which had previously been labeled mot. This behavior may have involved at least two mechanisms, which are not mutually exclusive: (*a*) The training with the candy attached a very strong reaching response to the specific cue on the box. This response generalized to the similar box and mediated the transfer of the name which was wrong for that box. (*b*) The training with candy so strengthened the tendency to respond to the specific cue on the box with the name of that box that the generalization of this name to the physically similar transfer box was able to override the previous training to give a different name to the transfer box.

In any event, the previous training in naming did not mediate secondary generalization for those *S*s who picked up the "incorrect" box.

As expected, the proportion of *S*s who failed to show secondary generalization on the basis of the verbal cues was relatively greater in the comparatively lower IQ groups. All the *S*s in the lowest group showed such failure. Nevertheless, the fact that some of the mongoloids were able to respond (albeit "incorrectly") on the basis of the correct verbal label suggests that the ability of this group of mental defectives to utilize verbal symbols in learning and performance may possibly be greatly improved by verbal training.

SUMMARY

The present study was concerned with the relationship between verbal cue-producing responses, intelligent behavior, and primary vs. secondary stimulus generalization. Under combined conditions of acquired distinctiveness and acquired equivalence of cues, groups of intellectually normal, high grade retarded, and middle grade retarded (mongoloid) boys were compared as to the utilization of learned verbal responses in the transfer of a manipulative response. As predicted, there was a significant relationship between IQ level and transfer score, indicating that the relatively higher IQ groups responded significantly more on the basis of secondary

stimulus generalization than did the relatively lower IQ groups.

REFERENCES

BIRGE, J. S. The role of verbal responses in transfer. Unpublished doctoral dissertation, Yale University, 1941.

BRACKBILL, Y. Experimental research with children in the Soviet Union: Report of a visit. *Amer. Psychologist*, 1960, **15**, 226–233.

CANTOR, G. S., & HOTTEL, J. V. Psychomotor learning in defectives as a function of verbal pretraining. *Psychol. Rec.*, 1957, **7**, 79–85.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

GOLU, M. The role of speech in the anal of simple visual and auditory signals children with the cerebroasthenic syndro In A. R. Luria (Ed.), *Problems of the hig nervous activity of normal and abnor children*. Vol. 1. Moscow: Academy Pedagogical Sciences of the RSFSR, 19 Pp. 250–266.

LINDQUIST, E. F. *Design and analysis experiments in psychology and educati* Boston: Houghton Mifflin, 1953.

MCNEMAR, Q. *Psychological statistics*. (2 ed.) New York: Wiley, 1955.

MILLER, N. E. Theory and experime relating psychoanalytic displacement stimulus-response generalization. *J. a norm. soc. Psychol.*, 1948, **43**, 155–178.

(Received August 8, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 403–408

DYNAMIC BEHAVIOR OF THE WARMTH SENSE ORGAN

E. EIJKMAN AND A. J. H. VENDRIK

*Department of Medical Physics, Roman Catholic University,
Nijmegen, The Netherlands*

Recent investigations, neurophysiological as well as psychophysical, have shown that temperature rather than temperature gradient should be considered as the adequate stimulus of the warmth sense organ (Hensel, 1950; Hensel & Zotterman, 1951a; Vendrik & Vos, 1958). However, it is clear that the time dependence of temperature also is important for the stimulation of this sense organ. Adaptation of the skin senses in general is very marked.

The term adaptation is often employed in a variety of ways. In this paper it will be used in the sense that the response to a constant physical stimulus diminishes with time. Hence, adaptation is a special type of dynamic behavior of this sytem. The response may be nervous activity measured electrophysiologically or sensation. A quantitative description of adaptation needs the measurement of the response to a physical stimulus with a well known time course. The measurement of the magnitude of a sensation, however, gives large difficulties. Therefore in psychophysical experiments the magnitudes of physical stimuli with various time courses are measured which give rise to a threshold sensation or another sensation of constant magnitude. These results allow the dynamic behavior to be calculated.

The comparison of electrophysiological and psychophysical data is very important. It allows one to interpret the electrophysiological responses in terms of their functional significance and to explain the psychophysically determined results. But this comparison is only possible when both responses are measured quantitatively.

Electrophysiological experiments on the sensory nerves of the temperature sense organs of the cat have revealed that the adaptation of the cold and warmth receptors in this animal is very pronounced but not complete (Hensel, 1953). The time-constant of this receptor adaptation has been measured. The adaptation of the sensation in man, however, is complete if the applied change in temperature is not too large. This points to another adaptation originating more centrally than the one of the receptors.

In a previous paper by Vendrik and Vos (1958) it was shown that adaptation of warmth sensation becomes very marked for stimulus durations greater than a few seconds. This adaptation does not agree quantitatively with the receptor adaptation measured electrophysiologically in the cat. The experiments by Vendrik and Vos (1958), however, were not designed for the investigation of this problem. The main limitation was that the durations of the stimuli could not be made short enough. Therefore the threshold measurements of the warmth sensation described in this paper were carried out with the aim to obtain more quantitative information on the adaptation, i.e., on the dynamic behavior, of this sense organ.

METHOD

For testing dynamic behavior it is necessary to use stimuli of quantitatively well known time course. Moreover, it was desirable to use stimuli of considerable shorter duration than 1 sec.

The dependence of temperature with time at the receptor layer of the skin is much more accurately known when microwave stimulation is applied than when infrared radiation or thermodes are used. The reason is that the absorption coefficient of microwave radiation of suitable wavelength in water-like tissues is so small that the skin above the fat layer is practically uniformly raised in temperature. Consequently heat conduction during short exposures does not play an important role. With infrared radiation or thermodes, by contrast, the rate of change of temperature in different skin layers is largely determined by heat conduction.

In present experiments microwave radiation with a free space wavelength of 10 cm. has been used to irradiate the inner side of the forearm. The arm was gently pressed against the aperture of a rectangular wave guide with a cross section of 7.2 × 3.6 cm^2. Also experiments with cross sections of 2.5 × 7.2 cm.^2 and of 6.0 × 7.2 cm.^2 were done. Greater variations in the cross-sectional area of the aperture of the waveguide cause reflections of the microwave to such an extent that a mismatch cannot easily be prevented. The equipment, its calibration and its use have been previously described (Vendrik & Vos, 1958).

When the arm is exposed to microwave radiation, a constant energy per unit of time is transmitted into the arm. During the time that heat loss to surrounding tissues is negligible, the temperature rise will be linear with time. From measured temperature rises on the surface of the skin, as well as from calculations, it appears that increase of temperature is linear within about 10% when exposure times do not exceed 7 sec. (Eijkman, 1959).

The experiments consist of measurements of threshold energy at different exposure times. The threshold measurements have been made using the yes-no procedure. The *S* was asked to make a decision whether or not a stimulus was presented to him during an indicated interval of time. A judgment on the existence of a stimulus is based on a neural activity that is probably fluctuating. To make a choice out of a fluctuating activity is in fact a problem of detection, recently investigated by Tanner and Swets (1954).

It is important to keep the *S*s uncertain about the fact whether a stimulus really exists. The experiments were conducted in such a way that *S* could freely decide when he wanted to make an observation but leaving him in doubt about the occurrence of a "zero" stimulus or a "real" stimulus. All series consisted of 10 trials, which by means of a programing switch, were composed of 5 stimuli of an announced strength and 5 "zero" stimuli in random sequence. Mostly well-trained *S*s were tested, who thus knew perfectly well, what a particular strength of stimulus meant. While *S* under test was observing the stimuli of the preset program, which was, of course, unknown to him, he wrote down the program he experienced. The possibility is now included, that a record of sensation is made, while no stimulus (i.e., a "zero" stimulus) was present, a so-called false positive. This gives the opportunity to describe the reliability of a normal report of sensation while a stimulus is present. The probability of perception is defined as the probability of report of sensation minus the probability of a false positive. This definition is chosen because this gives the best correction we can make at the moment for the variations in the observer's attitude. If one adheres to threshold theory, this definition gives the same result as the well-known definition used in threshold theory as regards the course of the curves describing the dynamic behavior of this organ. If one thinks, as we do, that detection theory is a better model, evaluation of the probability of obtaining a false positive can only be given if the distribution and the bandwidth of the noise is known. This knowledge, however, is lacking. Although not ideal the above mentioned definition provides a good correction for the shift in the observer's detection level.

The curve presenting the probability of perception versus the intensity of the stimulus is S shaped. The intensities are chosen in the region of the greatest slope of the S curve. Every experimental probability curve includes four intensities with the same exposure times; at each intensity six or seven series of 10 trials were performed. Eight different exposure times were used ranging from 0.1 to 10 sec. The *S*s were tested during about 3 weeks successively.

The threshold finally is defined as the energy that yields a probability of perception of .85, this being the percentage that could be determined most accurately. Intensities as well as exposure times were controlled electronically.

RESULTS

An example of a measured "probability of perception" curve is shown in Fig. 1. The stimulus which is

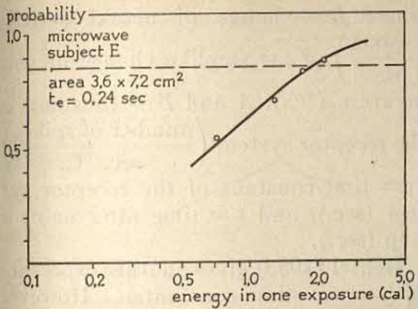


FIG. 1. Probability of perception versus total energy applied in one exposure of constant duration t_e. (The probability is corrected for false positives. The probability value 0.85 renders the threshold energy.)

plotted along the abscissa is the total energy given in one exposure. This quantity is in good approximation proportional to the temperature increase of the receptors at the end of the exposure. From these curves the thresholds defined as the energy necessary for a probability of perception of .85 are obtained. As all experimental probability curves showed the same shape, a standard curve was fitted to the measured probabilities. This could be done with a standard error of 15%; thresholds could thus be determined with an accuracy of 15%.

In Fig. 2 the thresholds of 2 well trained *S*s are plotted as a function of the exposure time. Similar threshold measurements with reduced and increased area of irradiation (2.5×7.6 cm.² and 6.0 × 7.6 cm.²) yield the same dependence on exposure time as the one shown in Fig. 2.

The characteristic features of the measured curves are: (*a*) Threshold energy increases initially with exposure times up to about 1–2 sec., which means that with a steeper rise of temperature a threshold sensation is obtained at a smaller increase of temperature. (*b*) For exposure

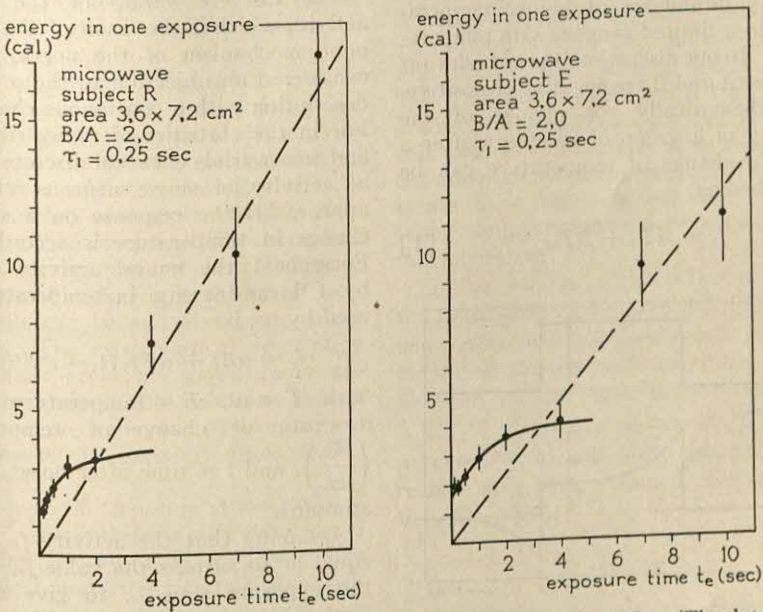


FIG. 2. The threshold energy as a function of the exposure time for 2 *S*s. (The dots represent experimentally determined values of threshold energy with their standard errors indicated by vertical lines; the solid line represents the theoretical curve of peripheral adaptation; and the broken line represents the theoretical curve of central adaptation.)

times larger than about 1 sec. the curve tends to run more horizontally. Threshold temperature is then much less dependent on the rate of change of temperature. (*c*) Finally, threshold energy increases proportionally with exposure time, the threshold rate of temperature change being constant.

DISCUSSION

In the first place, one can think of the dynamic behavior of the receptors as a possible explanation of the experimental results. Hensel and Zotterman (1951b) have investigated electrophysiologically the dynamic behavior of the cold and warmth receptors in the cat. After a sudden change of temperature, the temperature fibres respond with an initially considerable change of activity (overshoot) dropping to a new stationary value (Fig. 3). The quantities A and B (Fig. 3) of the temperature receptors mostly found by Hensel and Zotterman have negative values. This means that activity diminishes at higher temperatures in a limited range of skin temperatures. In our discussion it is not relevant whether A and B are negative or positive.

Mathematically the change of the activity of a group of receptors after a steplike change of temperature can be described as:

$$f_r = AT_s + BT_s e^{-t/\tau_1} \qquad [1]$$

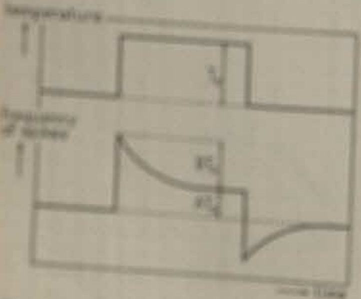


FIG. 3. Dynamic behavior of temperature fibres in cat and dog found in electrophysiological studies by Hensel and Zotterman (1951b).

where f_r = change of neural activity $\left(\frac{\text{spikes}}{\text{sec.}}\right)$, T_s = steplike change of temperature (°C.), A and B = constants of the receptor system $\left(\frac{\text{number of spikes}}{\text{sec. °C.}}\right)$, τ_1 = time-constant of the receptor system (sec.), and t = time after onset of step (sec.).

Hensel (1953) gives another expression with two time-constants. However, the influence of the second time-constant (0.01 sec.) is quantitatively negligible, and we do not think that experimental evidence thus far justifies the assumption of this second time-constant.

As will be seen, the above formula describes neural activity changing with regard to an already existing activity, which is not taken into account. In other words activity is measured counting from a physiological "zero." Naturally, the assumption is made, that the receptor system is working linearly for the small temperature rises needed to measure thresholds.

The discrete values of the neural activity which is the result of the all-or-none mechanism of the nerve, is not considered an objection to the proposed description with a continuous character. For in the statistics of many receptors and many trials eventual discrete values of activity of single neurons will disappear. If the response on a steplike change in temperature is according to Formula 1 the neural activity caused by a linear increase in temperature, as used by us, is:

$$f_r = aAt + aB\tau_1(1 - e^{-t/\tau_1}) \qquad [2]$$

with $T = at$, T = temperature (°C.), a = rate of change of temperature $\left(\frac{\text{°C.}}{\text{sec.}}\right)$, and t = time after onset of the stimulus.

Assuming that the activity f_r has to equal or to surpass the value f_0 during the exposure time t_0 to give rise to perception, the following expression is obtained for the threshold as a function of exposure time:

$$f_0 = aAt_0 + aB\tau_1(1 - e^{-t_0/\tau_1}) \qquad [3]$$

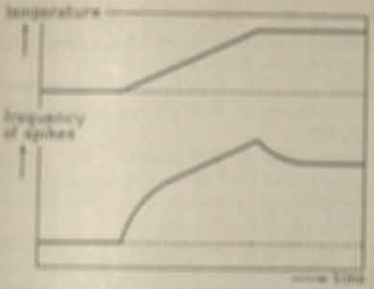


FIG. 4. Response of temperature fibres to a linearly changing temperature. (The curve is a consequence of the response shown in Fig. 3.)

This formula, which is suggested by the electrophysiological measurements, appears to describe the experimental results for exposure times smaller than about 3 sec. very well. The best fit to the experimental results is obtained by taking $B/A = 2.0$ and $\tau_1 = 0.25$ sec. (see Fig. 2). The same quantities of all experiments (3 Ss and three different areas) fall in the range $B/A = 2$-4 and $\tau_1 = 0.2$-0.4 sec. In electrophysiological experiments the cold receptors in the tongue of the cat appeared to have a time-constant in the range 0.3-2.2 sec. (Hensel, 1953). Records of the activity of warmth fibres show a dynamic response that might be described with a time-constant somewhat smaller than found for cold receptors. So the values of τ_1 in our experiments are in the range of comparable electrophysiological data. The same can be said about the values of B/A. The mathematical description of dynamic properties given above appears to be applicable to electrophysiological data and to psychophysical experiments with exposure times smaller than 3 sec.

However, if exposure times exceed 2-4 sec., the experimentally determined threshold energy deviates from the theoretical model described above. This means that for exposure times larger than 2-4 sec. the rate of change of temperature determines the perceptibility. If the peripheral adaptation model is accepted which is based on the electrophysiological studies and confirmed by our psychophysical results a nonperipheral adaptation must be assumed in addition which deals with the stationary activity of the receptors. Such an adaptation which we will call central adaptation, has been shown to exist. Landgren (1957) measured the cortical receptors of cold impulses from the tongue of the cat. His registrations show a marked adaptation. But as the corresponding activity in the peripheral nerves was not measured a quantitative relationship between peripheral and cortical activity is not known. Electrophysiological data of central adaptation which can be compared with our psychophysical results are lacking.

A quantitative description, therefore, can only be based on our own experimental results. It is perhaps plausible to assume that the response of the central nervous system to a constant input, i.e., peripheral nervous activity, decreases exponentially with time. It appears that the experimental curve is not fitted particularly well by this model. A better agreement is reached when the assumption is made that central adaptation occurs after a retardation time. This retardation time turns out to be 2-4 sec. More definite conclusions about this central adaptation must wait till the adequate electrophysiological experiments have been carried out. These experiments are now in preparation in this laboratory.

Similar studies about adaptation of the touch sense have been performed by us. Also electrophysiological experiments of the response of touch nerve fibres in the cat have been performed. A very satisfactory agreement between psychophysical and electrophysiological results have been obtained (Eijkman & Vendrik, 1960).

SUMMARY

Psychophysical experiments are described, which were carried out with the aim to investigate the adaptation of the warmth sensation. Microwave radiation was used as a warmth stimulus, giving a relatively

accurate knowledge of the time course of the temperature in the deeper layers of the skin.

The dependence of the threshold of the warmth sense on the time of exposure shows adaptational effects, which are in agreement with electrophysiological evidence. Hence, for exposure times up to 2–4 sec. the experimental results are described by a receptor system having a time-constant of about 0.3 sec.

Exposure times exceeding 2–4 sec. display another adaptation, which is believed to be of central origin. The obtained curves suggest that after the onset of the stimulus this central adaptation starts to diminish the effect of peripheral inflow after a retardation time of a few seconds.

REFERENCES

EIJKMAN, E. Adaptation of the senses of temperature and touch. Unpublished doctoral dissertation, Roman Catholic University, Nijmegen, The Netherlands, 1959.

EIJKMAN, E., & VENDRIK, A. J. H. Dynamics of the vibration sense at low frequency. *J. Acoust. Soc. Amer.*, 1960, **32**, 1134–1139.

HENSEL, H. Temperaturempfindung und intracutane Wärmebewegung. *Pflüg. Arch. ges. Physiol.*, 1950, **252**, 165–215.

HENSEL, H. The time factor in thermoreceptor excitation. *Acta physiol., Scand.*, 1953, **29**, 109.

HENSEL, H., & ZOTTERMAN, Y. Action-potential of cold fibres and intracutaneous temperature gradient. *J. Neurophysiol.*, 1951, **12**, 377–385. (a)

HENSEL, H., & ZOTTERMAN, Y. Quantitative Beziehungen zwischen der Entladung einzelner Kältefasern und der Temperatur. *Acta physiol., Scand.*, 1951, **23**, 291–319. (b)

LANDGREN, S. Cortical reception of cold impulses from the tongue of the cat. *Acta physiol., Scand.*, 1957, **40**, 202–209.

TANNER, W. P., & SWETS, J. A. A decision making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

VENDRIK, A. J. H., & VOS, J. J. Comparison of the stimulation of the warmth-sense organ by microwave and infrared. *J. appl. Physiol.*, 1958, **13**, 435–444.

(Received August 8, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 409–415

STRENGTH OF A GENERALIZED CONDITIONED REINFORCER AS A FUNCTION OF VARIABILITY OF REWARD[1]

RICHARD A. WUNDERLICH[2]

John Hopkins University

Recently Skinner (1957) has proposed that a stimulus which has been associated with more than one primary reinforcer will acquire the properties of a generalized conditioned reinforcer (GCR). A GCR, as proposed by Skinner, is a reinforcing state of affairs whose "reinforcing properties" are effective over a wide range of situations. It differs from a simple conditioned reinforcer (SCR) which controls behavior only under appropriate conditions of deprivation in not being dependent upon momentary deprivational conditions. Also, a GCR is assumed to have greater potential reward value than a SCR. Finally, Deese (1952) has pointed out that a habit based on a GCR should be more resistant to extinction than one based on a single reward.

Only a few studies (e.g., Porter & Miller, 1957; Wike & Barrientos, 1958; Wike & McNamara, 1955) have investigated the properties of a GCR, and these have been concerned primarily with the role that different drive conditions might have in the development of a GCR. For example, Wike and McNamara (1958) paired the stimuli of one goalbox at the end of an alley with wet-mash reinforcement of hunger and thirst drives in rats and the stimuli of another goalbox with wet-mash reinforcement of a hunger drive during the first phase of an experiment. Later, when confronted with both goalboxes in a T maze test, *S*s preferred the goalbox previously associated with reinforcement under both drives.

While the present literature suggests that conditions of drive are important variables in the determination of a GCR, studies examining other properties, such as the nature of the reinforcing situation itself, are lacking. The present experiment investigates the effects of varying (*a*) the type of reinforcement (food, water, or both) and (*b*) the distribution of the types of reinforcement during training trials, on the strength of a GCR as measured by resistance to extinction of a running response. Relative preference for the food and water rewards and spontaneous recovery were also used as supplementary measures in the study. Reward preference was used as a measure of reinforcement "strength"; spontaneous recovery, as another measure of a GCR.

METHOD

Subjects

Sixty male albino rats of the Sprague-Dawley strain obtained from Sprague-Dawley

---

[1] The experiment reported here was submitted as a dissertation in partial fulfillment of the requirements for the PhD degree at the Johns Hopkins University. The author wishes to express his sincere gratitude to Stewart H. Hulse who acted as faculty advisor to this dissertation. His suggestions and criticisms throughout all phases of the experiment were invaluable. The author is also grateful to Leon S. Otis for critically reading the manuscript.

[2] Now at the Department of Psychology and Psychiatry, Catholic University of America, Washington 17, D.C.

Inc., Madison, Wis., served as *S*s in the experiment. There were 15 *S*s in each of four groups. The *S*s were about 95 days old at the beginning of conditioning.

Apparatus

The apparatus was similar to that described by Hulse (1958). It consisted of a straight alley with start and goalboxes. The overall inner dimensions of the alley were 48 × 4 × 9 in.; the startbox, 7 × 6 × 9 in.; the goalbox, 12 × 8 × 9 in. The start and goalboxes were covered with hinged glass lids while the alley was covered with a hinged, screened frame. The interior of the startbox was white; that of the alley, flat black. The interior of the goalbox was also black with the exception of the back wall which was white. Two reward cups approximately 1.75 in. in diameter were mounted on the goalbox floor near the back wall. One reward cup was used for food reinforcement; the other, for water reinforcement.

Manually operated guillotine doors separated the start and goalboxes from the alley. Raising the door of the startbox started a Standard Electric timer. The timer was stopped when the weight of *S* depressed a floor panel in the goalbox. The relay circuits and timers were enclosed in a sound-damping box.

A light-gray box 12 × 10 × 10 in. was used to confine *S*s during a 15-sec. intertrial interval. The interval was measured by a timer which was started as *S* was placed in the box.

Experimental Design

Four groups of *S*s were run under both hunger and thirst motivation. The groups differed only in the type of reward that they received during conditioning trials. Group F received a food reward on all trials; Group W received a water reward on all trials; Group F + W received both food and water on all trials; and Group F/W received food randomly on half the trials and water on the remaining half. Three *S*s of each group were run during each of five replications of the experiment. One week separated the beginning of training trials for successive replications.

Procedure

Taming.—Upon arrival from the supplier, all *S*s were permitted ad libitum food and water for a period of 3 days. A 16-day period of taming followed. During the first 14 days, a deprivation schedule, which permitted an average daily food ration of 8 gm. of Purina chow and 13 cc. water, was instituted. Each day *S*s were taken from their housing cage and successively placed in groups of 6 in four different boxes. In the first box, they found one-half of their daily ration of food, in the second box, one-half their daily ration of water, in the third box the remaining half of their food, and in the fourth box, the remaining half of their water.

A 2-day exploration period followed in which *S*s in groups of 3 were permitted to explore the alley for two periods each day. Each *S* of Group F received approximately 0.50 gm. of dry food (0.045 gm/pellet)[3] in the goalbox during each of the two exploration periods. Each *S* of Group W received 0.50 cc. water during each of the exploration periods, while *S*s of Group F + W received both 0.25 gm. food and 0.25 cc. water on each period. Group F/W received 0.50 cc. water on the first period and 0.50 gm. food on the second period of the first exploration day; this order was reversed on the second exploration day. As previously mentioned, food rewards were always found in one cup in the goalbox while water rewards were found in the second cup.

Supplementary rations were given immediately after the exploration periods. These rations brought the average daily food consumption to 8 gm. and the average daily water ingestion to 13 cc. per *S*.

Conditioning.—All *S*s received 50 conditioning trials, 10 trials on each of 5 consecutive days. The intertrial intervals were 15 sec. Each *S* of Group F received two dry 0.045-gm. food pellets per trial as a reward. Each *S* of Group W received two drops of water (0.10 cc.) per trial as a reward. Each *S* of Group F + W received one food pellet (0.045 gm.) and one drop of water (0.05 cc.) per trial. Each *S* of Group F/W received as a reward either two food pellets or two drops of water on any one trial. The sequence of food or water for Group F/W was randomly determined with the restrictions that half the trials were water reinforced, that the initial trial of each day was water reinforced, and that no single kind of reinforcement occurred more than three times in succession. All *S*s of this group followed one of three reinforcement schedules.

As in the exploration periods, supplementary rations were given approximately 1 hr. after each daily session. These rations

---

[3] Obtained from P. J. Noyes Co., Lancaster, New Hampshire.

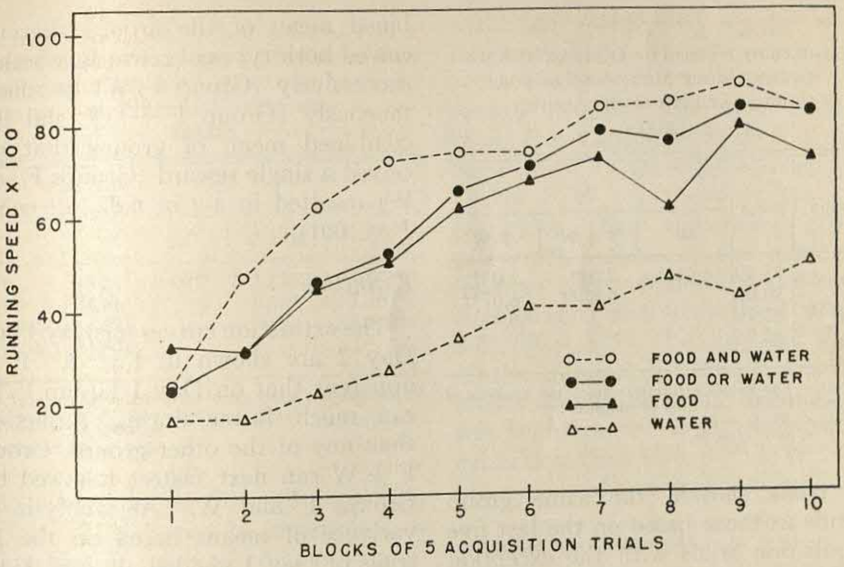


FIG. 1. Running speeds during acquisition.

brought the total daily food and water consumption to 8 gm. of food and 13 cc. of water. Supplementary rations were given to each *S* individually in separate cages

Extinction.—Following conditioning, 20 extinction trials were given. Ten extinction trials were given on the first day and 10 trials on the following day. During extinction, *S*s were permitted to remain in the empty goalbox for 5 sec. on each trial. The interval between extinction trials was 15 sec.

Performance Measures

During conditioning and extinction, the total running time per trial was recorded to the nearest .1 sec. Total running time was defined as the time from the raising of the startbox door to the time *S* entered the goalbox and depressed the floor panel. These measures were converted to reciprocals and multiplied by 100 for purposes of statistical treatment.

As a second measure of performance, spontaneous recovery occurring between the 2 extinction days was examined. To check for the statistical significance of spontaneous recovery, means of the reciprocal scores for the last trial of Day 1 of extinction were compared with means for the first trial of Day 2 of extinction.

Finally, the frequency with which the 15 *S*s of Group F + W consumed food before water in the goalbox during the 50 training trials was recorded. This "preference" measure was suggested when it was noted during taming that, under the experimental conditions used, *S*s showed less hesitancy in consuming dry food than in consuming water.

RESULTS

Conditioning

The acquisition data based on group means for blocks of five trials are shown in Fig. 1. An analysis of variance based on means for the last five conditioning trials yielded an *F* of 20.81 ($df = 3/56$, $P < .001$). Table 1 shows the results of *t* tests on the means for the last five conditioning trials. It is evident from the results of this analysis that Groups F/W and F + W ran significantly faster than Groups F and W but did not differ from each other in running speed.

An analysis of variance was also performed for the entire 50 acquisition trials. The resulting *F* of 25.50 was again significant ($df=3/56$, $P<.001$). The *t* tests between means based on

TABLE 1

RESULTS OF t TESTS OF DIFFERENCES BETWEEN GROUP MEAN SPEEDS FOR THE LAST FIVE CONDITIONING TRIALS

Group	Mean	t		
		W	F + W	F/W
F	73.49	4.64**	2.04*	2.03*
W	50.00		6.68**	6.67**
F + W	83.79			0.01
F/W	83.76			

Note.—The *within* error term from the analysis of variance (182.88) was used to compute t.
* $P < .05$, $df = 56$.
** $P < .001$, $df = 56$.

50 trials showed the same group trends as those based on the last five acquisition trials with the exception that Group F was not significantly different from Group F/W. Thus, although Fig. 1 suggests that Group F + W ran faster than Group F/W during training, the difference between these groups over the 50 acquisition trials was not significant.

Finally, a t test between the combined mean of the groups that received both types of rewards, whether successively (Group F/W) or simultaneously (Group F + W), and the combined mean of groups that received a single reward (Groups F and W) resulted in a t of 6.82 ($df = 56$, $P < .001$).

Extinction

The extinction curves for Day 1 and Day 2 are shown in Fig. 2. It is apparent that on Day 1 Group F/W ran much faster during extinction than any of the other groups, Group F + W ran next fastest followed by Groups F and W. An analysis of variance of means based on the 10 trials of Day 1 yielded an F of 23.19 ($df = 3/56$, $P < .001$). The results of t tests on differences between these means are shown in Table 2. All means are significantly different from each other.

An F of 10.45 ($df = 3/55$, $P < .001$) resulted from an analysis of covariance based on the last five condition-

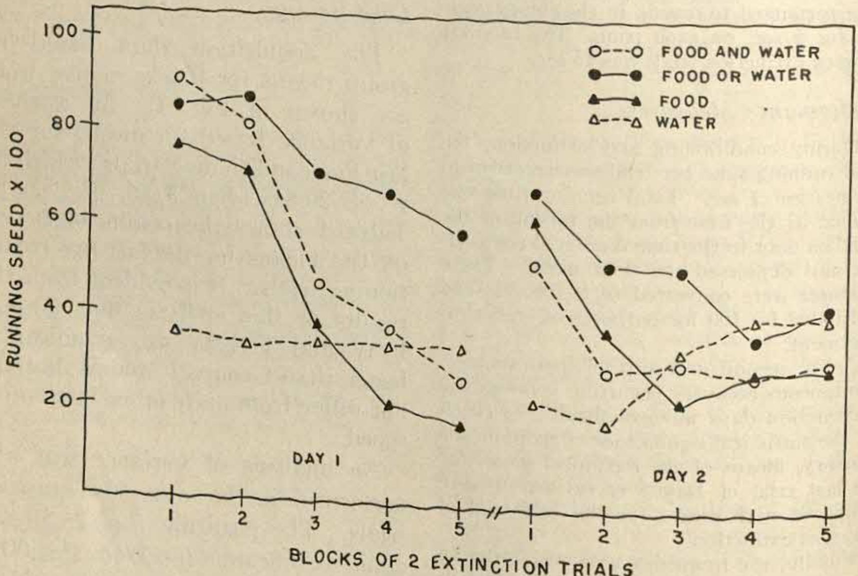


FIG. 2. Running speeds during extinction.

TABLE 2

Results of t Tests between Group Differences in Mean Speeds during the First 10 Extinction Trials

Group	Mean	t		
		W	F + W	F/W
F	45.51	2.09*	2.29*	5.67**
W	34.80		4.37**	7.76**
F + W	57.29			3.38**
F/W	74.67			

Note.—The *within* error term from the analysis of variance (188.77) was used to compute t.
* $P < .05$, $df = 56$.
** $P < .01$, $df = 56$.

ing trials and the first 10 extinction trials. This F may be interpreted as showing an overall between-group difference after accounting for the differences between groups at the end of conditioning.

Table 3 presents the means and the t's between group means based on the first 10 extinction trials after the means have been adjusted by the covariance technique. As before, Group F/W shows significantly faster running during extinction than any of the other groups. On the other hand, the greater resistance to extinction of Groups F and F + W, as compared with Group W, that was found with the unadjusted means is no longer evident.

TABLE 3

Results of t Tests between Adjusted Group Differences in Mean Speeds during the First 10 Extinction Trials

Group	Adjusted Mean	t		
		W	F + W	F/W
F	45.25	0.40	1.61	5.08*
W	42.99		1.91	4.29*
F + W	53.32			3.59*
F/W	70.71			

Note.—Adjusted standard errors of the mean difference were used to compute t.
* $P < .001$, $df = 55$.

An analysis of variance of the second 10 extinction trials resulted in an F of 4.86 ($df = 3/56$, $P < .01$). Subsequent t tests again showed that Group F/W was more resistant to extinction than the other groups ($P < .05$). None of the other differences were significant, however. A covariance analysis yielded an F of 3.62 ($df = 3/55$, $P < .05$) and t tests of the differences between adjusted means showed only that Group F/W was significantly more resistant to extinction than Group F + W. The other comparisons were not significant.

Spontaneous Recovery

As a measure of spontaneous recovery, an analysis of variance was performed on the difference between running speeds of the last trial of Day 1 and the first trial of Day 2 of extinction. The resulting F of 26.22 was significant ($df = 1/118$, $P < .001$). Only Groups F and F + W showed significant spontaneous recovery. The t for Group F + W was 5.02 ($df = 118$, $P < .001$); for Group F the t was 3.98 ($df = 118$, $P < .001$).

Reward Preference

A t of 4.57 ($P < .001$) was found between the percentage of trials that *S*s of Group F + W consumed food before water (82%) and chance (50%).

Discussion

The principal finding of this experiment is that the *random distribution* of two types of rewards from trial to trial during training resulted in the clearest evidence of a GCR as measured by resistance to extinction. Thus, Group F/W ran much faster during extinction than either Group F + W, Group F, or Group W. Simply using the two rewards in the goalbox on *each* training trial, under the present conditions, was not sufficient to give any clear indication of a GCR. Thus, Group F + W was running at about the same speed as Groups F and W at the end of each extinction day (cf. Fig. 2) and was not significantly different from these groups after extinction means were adjusted by the analysis of covariance.

These findings can be explained if it is first noted that the food reward used in the experiment was, operationally, a much "stronger" reinforcer than the water reward. This conclusion follows from two things: (*a*) Group F ran much faster than Group W during training, and (*b*) Group F + W reliably "preferred" food over water as determined by first reward selection in the goalbox.[4] Further, a recent finding by Logan, Beier, and Kincaid (1956) showed that aperiodic reinforcement of a running response with nine food pellets on half the training trials and one pellet on the other half produced extinction performance comparable to that produced by giving partial reinforcement with nine pellets on half the trials and nothing on the other half. Both these groups were more resistant to extinction than a group given nine food pellets on all training trials. On the assumption that nine food pellets are a "stronger" reinforcer than one food pellet (Logan, Beier, & Ellis, 1955), these data suggest that varied strength of reinforcements during training acts like partial reinforcement in increasing resistance to extinction.

Generalizing from the data of Logan et al. (1955) to the present findings, it would be predicted that Group F/W would show greater resistance to extinction than the other three groups because it was trained with an operationally "strong" reinforcer on half the training trials and an operationally "weak" reinforcer on the other half. Groups F + W, F, and W, on the other hand, received a constant "strength" of reinforcement from trial to trial, and it would be predicted that they would extinguish relatively rapidly, as they did.

It must not be assumed that, in the present experiment, simple amounts or weights of food and water rewards determined reinforcement strength. Although, for example, Group F + W received only half the weight of food that Group F/W received on any one trial (both groups, however, received the same total weight over all 50 trials), the running speeds during training were not significantly different. Similarly, Groups F and W ran much slower than Group F + W although the latter group received only half as much food and water on any training trial as the former groups. Clearly some effect due to the combination of food and water as a reward, which was independent of their simple weights, accounted for the faster training running speed of Groups F + W and F/W.

Finally, the data from spontaneous recovery did not give any consistent indication of a GCR. These data can be accounted for by noting that the two groups which did show significant spontaneous recovery (Groups F + W and F) were the only ones to show any appreciable loss in running speed over the 10 trials of Day 1 of extinction. Consequently, they were the only ones which could demonstrate much of an increase in running speed on the first trial of Day 2 of extinction.

SUMMARY

Four groups of rats under hunger and thirst motivation were used to investigate Skinner's notion of a generalized conditioned reinforcer (GCR). The groups differed only in the type of reward (food, water, or both) and the distribution of the types of reward that they received during conditioning of a running response in a straight alley.

---

[4] Related to the preference of food over water was a finding that has practical significance. Under the drive conditions in this experiment, it was found that *S*s that are both hungry and thirsty will accept dry food pellets as reinforcement. Contrary to previous reports, *S*s in this experiment showed no hesitation in accepting dry food pellets in either the reward situation or in the acceptance of food before water in the daily feeding. The small size of the food reward may have been a factor here.

The results showed evidence of a GCR in that a group that received either food or water reinforcement randomly from trial to trial demonstrated greater resistance to extinction than groups receiving both food and water, only food, or only water from trial to trial.

In addition to the importance of the distribution of the types of reinforcement in the formation of a GCR, it was suggested that differences in reinforcement "strength" may serve as an additional variable of some importance in developing a GCR.

REFERENCES

DEESE, J. *The psychology of learning.* New York: McGraw-Hill, 1952.

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

LOGAN, F. A., BEIER, E. M., & ELLIS, R. A. Effect of varied reinforcement on speed of locomotion. *J. exp. Psychol.*, 1955, **49**, 260–266.

LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

PORTER, L. W., & MILLER, N. E. Training under two drives, alternately present, vs. training under a single drive. *J. exp. Psychol.*, 1957, **54**, 1–7.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1957.

WIKE, E. L., & BARRIENTOS, G. Secondary reinforcement and multiple drive reduction. *J. comp. physiol. Psychol.*, 1958, **51**, 640–643.

WIKE, E. L., & MCNAMARA, H. J. A quest for the generalized conditioned reinforcer. *Psychol. Rep.*, 1955, **1**, 83–91.

(Received August 10, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 4, 416–422

PSYCHOPHYSICAL JUDGMENTS OF PROBABILISTIC STIMULUS SEQUENCES[1]

WILLIAM SIMPSON[2] AND JAMES F. VOSS

College of Wooster

The results of simple probability learning experiments, such as light guessing in a two-choice situation, generally support the matching rule of stimulus probability and *S*'s response probability (Estes, 1959). These findings are in conflict with the results of a more complex experiment where six probabilistic verbal associations were concomitantly learned. In the latter experiment, it was found that probabilities from .60 to .90 yielded systematic overresponding of the respective stimulus probability, and the probabilities of .10 to .40 yielded corresponding systematic underresponding (Voss, Thompson, & Keegan, 1959). One suggestion of the latter paper was that the response tendencies are related to the probability discriminations of *S*s. The present paper reports two experiments designed to investigate probabilistic psychophysical judgments and suggests a relation of probabilistic discrimination and probability learning.

The psychophysical investigation of thresholds typically involves presentation of a stimulus event and subsequent judgment by *S* with respect to some aspect of the stimulus presentation. The present experiments extend psychophysical investigation to judgments of probabilistic events. The stimulus was a sequence of 50 light presentations, where one and only one of two lights was presented on each presentation. Stimulus probability was manipulated by variation in the relative number of each of the two lights in the sequence. Judgments made at the end of each series of 50 lights regarding the estimated occurrence of one light provided a response measure of estimated probability.

EXPERIMENT 1

Method

Two 7-w. light bulbs, labeled A and B, were mounted at an equal height on a upright board. The light sequences were presented by *E*'s manipulations of switches that were behind and below the stimulus light board and out of *S*s' view. A low volume electronic metronome was used to maintain a 1 light/sec rate.

Eighteen groups of *S*s were randomly assigned to a 9 × 2 factorial design. The nine probability conditions (PC) employed were: .10, .20, .30, .40, .50, .60, .70, .80, and .90. The light judged, A or B, was orthogonal to PC. The *N*/cell was 10 and groups of 10 were run whenever possible.

The stimulus *sequences* were presented by the method of single stimuli (Woodworth & Schlosberg, 1954) with the modification that each stimulus occurrence was a sequence of 50 light presentations. Use of this number of light presentations enabled the probability of one light to vary from 0 to 1.00 in steps of .02. Eleven probability values of randomly obtained light sequences were used at each PC. These values centered about the PC value in steps of .02; e.g., the probability values of the light judged at the .60 PC were .50, .52, .54, . . ., .66, .68, .70. Upon presentation of each light sequence, *S*s judged whether the respective light occurred "more than" or "less than" the particular PC value.

At each PC, four random *series* of the 11 probability values were employed. Thus each *S* made 44 judgments (11 sequences at the respective PC × 4 random series of these values).

---

[1] The research reported in this article was supported by Grant M-3531 of the National Institute of Mental Health.

[2] Now at the University of Wisconsin.

The *Ss* were students at the College of Wooster and were instructed as follows:

> This is an experiment in your ability to make accurate judgments regarding the frequency of occurrence of lights. You are to make the judgments on the paper handed out to you. Lights A and B will come on individually a given number of times. At the end of the sequence, I will say "Judge." This means that you are to check the word that indicates whether you think Light A B came on more than or less than per cent of the times I have presented the light. So if you feel Light A B came on more than —— per cent, check "more;" if you feel it came on less, check "less." You are to check "more" or "less" but not both. Please make no effort to count the lights. You are to observe the lights and write your judgment when I say "Judge." When you have checked the word, turn that page over and do not look at it again. Are there any questions?

Results

Figure 1 presents the percentage of "more than" responses as a function of PCs for the four series. Since one-half of the sequences judged at each PC were above and one-half below the respective PC, the Series 1 data indicate a systematic tendency to overestimate the sequences at PCs greater than .50 and underestimate the sequences at PCs below .50. On the subsequent three series, the judgments of the .60, .70, and .80 PC sequences remained above the 50% response level, and the sequences of the .10, .20, and .30 PCs yielded responses at approximately the 50% level.

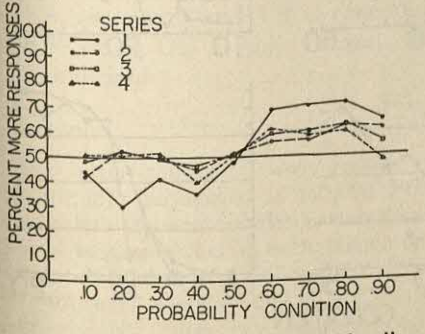


FIG. 1. Percentage of "more than" responses as a function of probability condition with Series as parameter.

An analysis of variance performed on the data of each PC revealed a significant Series source of variation at each PC except .30 and .50. The Light Judged (A or B) and Light Judged × Series interactions are not significant for any PC.

Since the data for most PCs indicated that the significant series effect is likely due to differences between Series 1 and the remaining three series, the data of Series 2, 3, and 4 were pooled for further analysis.

The Series 1 data were analyzed by a test of skew-symmetry in order to determine if the PC data were symmetric about the .50 PC. The frequency of "less than" responses was employed for the .10–.40 PCs, whereas the "more than" response frequency was used for the .60–.90 PCs. The analysis consisted of a 2 × 4 factorial design with the respective sources of variation of Above vs. Below the .50 PC and Amount Above and Below the .50 PC.

The Above vs. Below source of variation is significant ($F = 6.56$, $df = 1/152$). This finding indicates a significantly greater number of "more than" responses at PC above .50 than "less than" responses at PC below .50. The same analysis performed on the pooled data of Series 2, 3, and 4 again revealed a significantly greater number of "more than" responses at the .60–.90 PCs than the "less than" responses at the .10–.40 PCs ($F = 13.59$, $df = 1/152$). The Amount and Amount × Above vs. Below sources of variation are not significant in either the Series 1 or the pooled series 2–4 analyses.

Figure 2 presents psychophysical curves for the data of Series 1 and the pooled series 2–4. The .60–.90 PC pooled Series 2–4 data indicate a relatively large frequency of "more than" responses over all stimulus sequences, but an especially high frequency at stimulus sequences *below* the particular PC value (e.g., .70 in the .70 PC). The Series 1 data of the .60–.90 PCs indicate the same tendencies, with a generally greater response frequency at sequences below the respective PC values.

The .10–.40 PC Series 1 data indicate a corresponding tendency in that sequences *above* the PC value sequences tend to yield relatively fewer "more than" responses. The pooled Series 2–4 data for the .10–.40 PCs indicate, however, an increase in "more than" responses at sequences above the PC value sequences.

Figure 2 also reveals that the sequences at the particular PC values of .60–.90 were generally overestimated in both Series 1 and the pooled Series 2–4. The initial tendency in Series 1 of the .10–.40 at the PC value sequences was to underestimate; but in the pooled Series 2–4, judgments of sequences at the particular PC values varied.

Table 1 presents the PSE and DL values for Series 1 and the pooled Series 2–4. The PSE values for Series 1 of .60–.90 PCs reflect the asymmetric results of Fig. 2. Thus, since the responses of the stimulus sequences below the respective PC

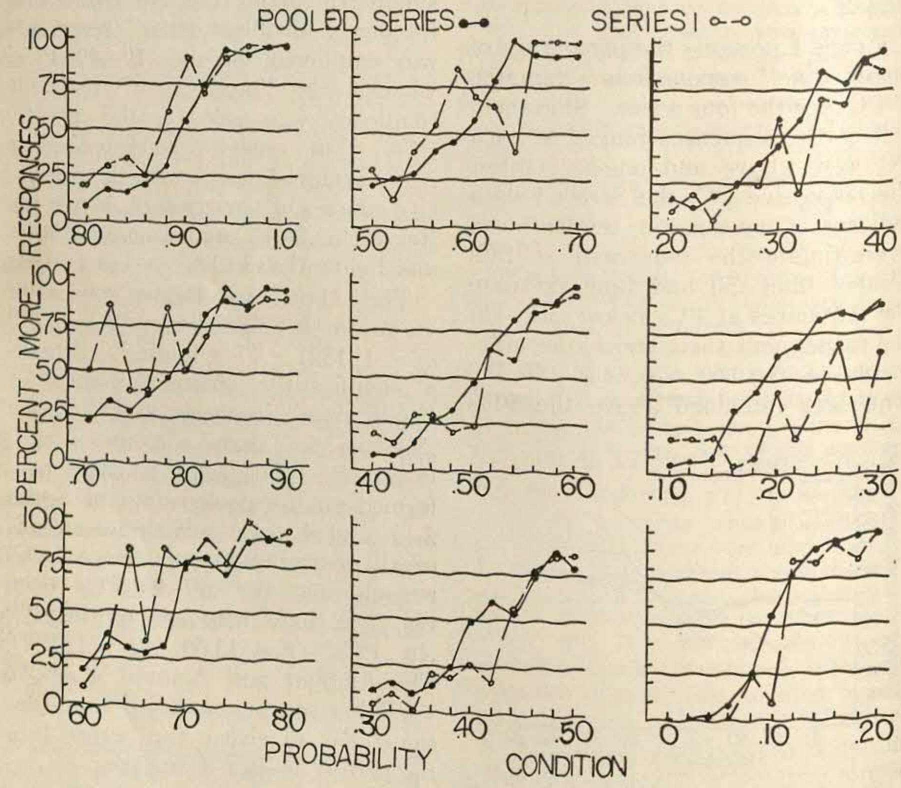


FIG. 2. Percentage of "more than" responses for the sequences of each PC with Series 1 and pooled Series 2–4 as parameter.

TABLE 1

PSE and DL Values for the Probability Conditions of Exp. 1 and 2

PC	Experiment 1				Experiment 2			
	Series 1		Pooled Series 2–4		Standard		Nonstandard	
	PSE	DL	PSE	DL	PSE	DL	PSE	DL
.90	.874	.048	.896	.027	.112	.014	.102	.021
.80	.800	—	.785	.045	.203	.023	.203	.046
.70	.666	.053	.688	.069	.308	.023	.302	.053
.60	.553	.067	.589	.049	.392	.022	.409	.054
.50	.513	.045	.502	.039	.498	.024	.511	.033
.40	.438	.049	.400	.041	.589	.037	.592	.039
.30	.300	.049	.307	.036	.691	.036	.694	.033
.20	.244	—	.187	.041	.795	.019	.793	.048
.10	.111	.019	.096	.020	.891	.019	.897	.022

sequences received a relatively large number of "more than" judgments, the 50% response point is found on the lower side of the PC value. Likewise, the PSE values of Series 1 for the .10–.40 PCs indicate a PSE above the PC values because of a relatively low level of "more than" judgments above the particular PC values. The pooled Series 2–4 PSE values generally remained on the same side of each PC as for the Series 1 .60–.90 PC values, but for then.10–.40 PCs, less consistency occurred.

The DL values generally indicate a reduction in variability on the pooled Series 2–4 compared to Series 1. Also, the restriction of variation in sequences near 0 and 1.0 is reflected by the small DL of the .10 and .90 PC conditions.

Discussion

The initial tendency to overestimate the stimulus sequences at .60–.90 PCs is especially associated with overestimation of sequences below each respective PC value. It should be noted that if all sequences at the .60–.90 PCs are overestimated, there is a greater possibility for an increase in "more than" responses at sequences below the PC value. The initial underestimation of the .10–.40 PCs is associated with fewer "more than" responses at the stimulus sequences above the respective PC. The Series 1 PSE values as well as the judgments made at the particular PC value sequences are in general agreement with the asymmetry of responses at the respective PC.

The pooled Series 2–4 data indicate that the initial tendencies diminish, although the .60–.90 overestimation response level persisted more than the .10–.40 underestimation tendency. Figure 2 reflects these differences in that the asymmetry of responses was maintained at the sequences of the .60–.90 pooled Series 2–4, whereas the .10–.40 pooled Series 2–4 data indicate more symmetric curves.

The four particular series at each PC and practice over the series are confounded in the significant series effect. The consistency over PC of the initial overestimation of .60–.90 PCs and the initial underestimation of .10–.40 PCs strongly suggests that although a certain degree of variability is associated with the intraseries differences, the Series 1 and the pooled Series 2–4 results are due largely to initial estimation tendencies.

EXPERIMENT 2

Method

Experiment 2 was designed to determine the generality of the findings of Exp. 1 under conditions of greater stimulus control and more standardized conditions. In particular, practice was extended by use of a 9 × 9 latin square design in which each group of *S*s served in all nine PCs. In addition, a standard stimulus was employed for one-half of the *S*s. The standard involved presentation of the particular PC value to be judged by showing *S*s light sequences of the PC value.

The procedure varied from Exp. 1 in that one randomly permuted 9 × 9 latin square was used for both the standard (S) and nonstandard (non-S) conditions. The Latin letters were the PC values of Exp. 1. Eight *S*s served in each row of the square.

Random sequences of 50 lights were presented at each PC value, as in Exp. 1, with the modification that instead of 11 probability values at each PC, 9 probability values were used which centered about the respective PC value in steps of .02. Furthermore, since each group served at each PC, the 9 probability values at each PC were permuted for the 9 groups (rows) by use of a 9 × 9 latin square. Thus, the 9 stimulus probability values within each PC were virtually counterbalanced over the 9 groups.

The same latin square was employed for the S group as for the non-S group. In addition, a standard sequence of lights, which always contained the PC value to be judged, was presented before each sequence of lights to be judged. In order to eliminate the possible effect of a particular standard sequence, nine such standard sequences were employed at each PC and constituted the Greek letters of a 9 × 9 greco-latin square.

The presentation rate and apparatus were the same as those of Exp. 1. Instructions were modified only to take into account the use of a standard and the changing PC conditions.

Results

Figure 3 presents the first PC judged by all groups in Exp. 1 and 2. Experiment 2 data of Fig. 3 were analyzed by the same test of skew-symmetry used in Exp. 1, with the modification that the S vs. non-S source of variation was included (2 × 2 × 4 factorial analysis). The significant effect of the S vs. non-S conditions ($F = 11.0$, $df = 1/112$), in conjunction with the lack of any significant S vs. non-S interaction, indicated that the non-S group made a greater number of responses than the S group ("more than" responses for .60–.90 PCs and "less than" for .10–.40 PCs). The only other significant source of variation is Amount of deviation from the .50 PC ($F = 3.5$, $df = 3/112$). However, this result only indicates that the .70–.30 PC conditions deviated more from the 50% level than the other three virtually equal Amount conditions.

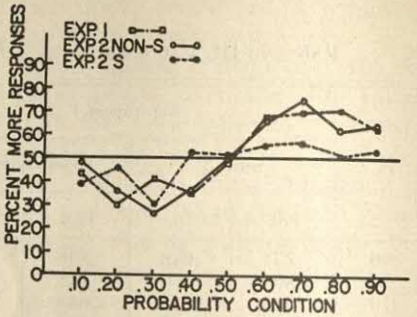


FIG. 3. Percentage of "more than" responses as a function of probability condition for the first judged probability condition of the standard and nonstandard conditions of Exp. 2 and for Exp. 1.

The latin square analysis performed on the more than response frequency data revealed a significant PC source of variation ($F = 2.34$, $df = 8/1408$), and a lack of significance of the S vs. non-S and S vs. non-S × PC sources of variation. The only significant source of variation in addition to PC is PC sequence ($F = 4.91$, $df = 8/8$). Inspection of the PC sequences of the latin square revealed that occurrence of the .70, .80, and .90 PCs early in the sequence produced a greater number of "more than" responses over the PC sequence, whereas occurrence of the .70, .80, and .90 PCs later in the sequence yielded fewer "more than" responses over the sequence.

The psychophysical curves for the S and non-S conditions were similar to the pooled Series 2–4 curves of Fig. 2. The major difference in the S and non-S curves is that the S condition apparently produced more uniform curves with more accurate extreme judgments than the non-S condition.

Table 1 presents the PSE and DL values for the S and non-S groups. The PSE values again generally occurred on the same side of the respective PC as found in Exp. 1. In addition, a slight tendency occurred for the PSE to be higher in the non-S condition. The DL results indicate that, especially with lower PCs, the standard reduced variability of judgments.

Discussion

The results of both experiments indicated an initial tendency to overestimate PC sequences above .50 and underestimate PC sequences below .50. Furthermore, Exp. 1 results indicated that the former tendencies were greater than the latter. The lack of Above-Below differences in Exp. 2 suggests that the relatively greater estimation of .60–.90 PC may be a function of the particular initial stimulus sequences employed.

Figure 1 and the significant Series effect of Exp. 1 suggest that with practice, the initial response levels tend to approach the 50% response level. The .60–.80 PCs, however, remained consistently above the 50% level and significantly different from the .10–.40 PCs underresponding. The practice effect (Ordinal Position) of Exp. 2 is confounded with PC sequences. Moreover, since sequence is significant, Exp. 2 suggests that the practice effects vary with PC sequences previously judged. Therefore, because of the use of only four series in Exp. 1 and because of the PC sequential effects in Exp. 2, no conclusions may be reached regarding long term practice effects. In particular, the amount of practice necessary to reduce the overestimation judgments of the .60–.80 PCs of Exp. 1 to the 50% response level, if such effects occur, cannot be specified.

The use of a standard stimulus apparently has the effect of making more accurate judgments early in practice and at the extreme sequence values of each PC. Moreover, the lack of a significant difference between the S and non-S conditions in the latin square analysis indicates that the differences of S and non-S are essentially eliminated by practice.

It should be noted that the use of 50 light presentations as a stimulus event implicitly contains a sequential factor not found in single stimulus events. The results of the present experiments therefore may be dependent to a presently unknown degree on the use of particular sequences of 50 lights.

One methodological problem of the present experiments is counting. At a 1 light/sec rate, there is the possibility of counting. However, *S*s were not provided with any suggestion or material to compute percentages, and judgments were made in an interval of only 5 sec. Moreover, it would be suspected that if *S* counted, he should have each judgment correct. Inspection of the data revealed that no *S* consistently or even often made all correct judgments at any PC or in any series.

The initial tendencies of the present results are somewhat in agreement with other studies on probability prediction and estimation. Neimark and Shuford (1959) found that in a two-choice situation, *S*s instructed to estimate and predict the occurrence of events estimated accurately, but overpredicted a compound event which involved simultaneous presentation of two samples of matrix elements. The present experiments however involved judgments after a sequence of stimuli, whereas Neimark and Shuford (1959) employed judgments after each light presentation, and Shuford (1959) studied judgments of simultaneous stimuli.

The discrepancy between the one-association situation (Estes, 1959) and the more complex six-association situation (Voss et al., 1959) may be due to the aspects of the stimuli to which *S*s respond. In the one-association case, *S*s likely respond primarily to sequential characteristics such as run lengths, etc. (Anderson, 1960; Nicks, 1959), whereas in the six-association paradigm, interference is produced between the stimuli and responses, and *S*s respond not to sequential characteristics, but primarily to frequencies of the events. Responding to frequencies thus produces results similar to the present experiments in that the probabilities above .50 yield overresponding and the probabilities below .50 produce underresponding.

Finally, the present results are relevant to decision making in that they present a functional relation of subjective probability as a function of objective probability in a rewardless, riskless situation. This function provides a baseline for comparison of the relation of subjective probability and objective probability under conditions of reward and risk (Edwards, 1954).

SUMMARY

Two experiments were reported on the ability of *S*s to judge the probability of occurrence of a given light. Sequences of 50 lights were presented where one of two lights was shown on each occurrence. Probability of a particular light in a 50-light sequence was from .10 to .90 in .10 interval steps. At each probability value, stimulus sequences were presented in random order in probability steps of .02, centered about the particular probability value.

The results indicated an initial systematic tendency of *S*s to overestimate probabilities from .60 to .90 and underestimate probabilities from .10 to .40. Experiment 1 indicated that the former tendency was greater than the latter, both initially and after practice. Experiment 2 did not provide information regarding practice except that sequence of PC employed is significant. Initial judgments were more accurate when the standard was used in Exp. 2. The results are discussed in relation to probability learning and decision making.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

EDWARDS, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380–417.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

NEIMARK, E. D., & SHUFORD, E. H. Comparison of predictions and estimates in a probability learning situation. *J. exp. Psychol.*, 1959, **57**, 294–298.

NICKS, D. C. Predictions of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

SHUFORD, E. H. A comparison of subjective probabilities for elementary and compound events. *U. N. C. Publ.*, 1959, No. 20.

VOSS, J. F., THOMPSON, C. P., & KEEGAN, J. H. Acquisition of probabilistic paired associates as a function of S-R$_1$, S-R$_2$ probability. *J. exp. Psychol.*, 1959, **58**, 390–399.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received August 11, 1960)

Journal of
Experimental Psychology

VOL. 62, NO. 5 NOVEMBER 1961

PERCEIVED ORDER IN DIFFERENT SENSE MODALITIES[1]

IRA J. HIRSH AND CARL E. SHERRICK, JR.

Central Institute for the Deaf, St. Louis, Missouri

Perceptual discrimination and recognition require, as different stimuli, patterns that are formed by intensitive and qualitative change over space or time. In vision, patterning in space appears to be of primary importance, while in hearing the important dimension for patterning appears to be time (Hirsh, 1952; Piéron, 1952). Discrimination of these patterns requires a certain resolving power in the observer. In visual space the resolving power that appears to be crucial is visual acuity or the minimal separation between two lines or dots for perception of two, as it is also for space laid out on the skin. Spatial resolving power has also been studied in hearing, where it takes the form of the minimum separable angle between two sound sources for differential localization.

For patterns that are articulated in time, such as movement in vision or touch, and melody in hearing, a temporal resolving power is required. Its measurement, however, is complicated by the unidimensional nature of time. Figures and forms in visual or tactual space may be recognized and retain certain form characteristics even when they are reversed, but temporal patterns are quite different when they are reversed in time. Temporal resolving power must, therefore, be broken down into at least two measures: one, analogous to spatial acuity, is the minimum time that can separate two events so that they are just barely nonsimultaneous or successive; the other, given that two events are perceived as successive, requires a judgment of temporal order to be made.

The requirements for separating questions of successiveness from those of successive order have not only to do with experimental method, but also with the kinds of stimuli used. To investigate the limits of the ability to distinguish successive from simultaneous events, we need only a pair or a series of events about which we can ask *O* whether there were one or two or whether the series consisted of separate discrete events or one continuous event. When we investigate the ability to recognize temporal order, on the other hand, the two events in question must be distinctive so that *O* has some way of labeling them differently and then of using these labels to identify which occurred first or second. Two identical stimuli that follow each other in time may be used to distinguish successiveness from simultaneity, but the question of successive order is not appropriate unless the two stimuli can be identified separately.

---

[1] This work was supported by a grant (G-4457) from the National Science Foundation to the Central Institute for the Deaf.

Simultaneity vs. successiveness.—One of the most extensive investigations having to do with successiveness was reported by Exner in 1875. He set for himself the task of determining how close together in time two stimuli may be and yet be perceived as not simultaneous (*ungleichzeitig*). Unlike more recent *E*s, Exner (1875) explored this temporal acuity in different sense modalities. In vision, he found that two brief flashes of light appeared to be nonsimultaneous when they were separated by 44 msec. or more. He reported certain variations in the result depending upon whether the stimulation was central or peripheral and upon whether both flashes stimulated the same place or different places. There were some conditions in which he reported apparent movement and in these cases he found that movement could be distinguished from simultaneity when the temporal separations were between 15 and 20 msec. He regarded movement, however, as a kind of artifact and relied more upon his 44-msec. value as descriptive of the temporal resolving power of the eye.

In hearing, using the successive spokes of a Savart wheel, he determined that the minimum temporal separation between successive clicks that would give rise to a judgment of successiveness was about 2 msec. The resolving power of the ear was clearly superior. When two clicks were delivered separately, one to each ear, the minimum temporal separation for successiveness was as large as 64 msec. Such a large difference probably has less to do with the two ears than with different experimental tasks, as will be noted presently.

Exner (1875) also reported some experiments on the minimum temporal separation between stimuli, one of which was presented to one sense modality and the other to another. The values are more like the large times of the visual temporal resolution or the one associated with stimulating the two ears separately, and furthermore they are not symmetrical; that is, in cases of intermodal stimulation, the value of temporal separation required for successiveness depended upon which of the two particular sense modalities was stimulated first.

Now Exner (1875) pointed out that there are four conditions under which successiveness versus simultaneity may be studied: (*a*) the stimuli could affect the same sensory elements, (*b*) they could affect different sensory elements in the same sense modality, (*c*) they could affect corresponding elements in paired sense organs, and (*d*) they could affect different sense modalities. More recently, Piéron (1952) and Fraisse (1957) have pointed out that the perception of simultaneity is easiest to obtain when stimuli occur at the same place and are presented to the same sense modality. Indeed, Fraisse (1957, p. 110) has indicated an order of difficulty among three classes of conditions which correspond to the first, second, and fourth conditions of Exner given above.

Particular values for the temporal acuity of different sense modalities based upon the distinction between successiveness and simultaneity have been given by many authors. Exner's (1875) values for vision and for hearing agree reasonably well with more recent data on visual flicker (Bartley, 1958, pp. 120–121), and on auditory successiveness (Miller & Taylor, 1948; Wallach, Newman, & Rosenzweig, 1949). Values of temporal acuity for the skin are less well established (Geldard, 1940). There is some evidence that different *E*s have used quite dissimilar criteria for tactual fusion. It is clear that this simple temporal acuity varies with the sense modality used and may indeed be ascribed to the resolving power of peripheral mechanisms.

As Exner's (1875) values for simple successiveness have been reported in secondary sources through the years (James, 1890, pp. 614–615; Ladd, 1892, pp. 364–366; and Troland, 1929, p. 399), they have been lumped together with the results of his other experiments involving stimuli presented to the two ears separately, to two parts of the retina, and to two different sense modalities. Fraisse (1957) comments that these latter cases of heterogeneous stimulation represent conditions in which perception of simultaneity is difficult to obtain. Indeed, although Exner (1875) is not very explicit about the particular instructions that he gave to his *O*s, it is doubtful that he asked them exactly the same questions in cases of homogeneous and heterogeneous stimulation. He does comment that he did not note a single case in which the time difference required for the perception of nonsimultaneity was different from that required to perceive which of the two stimuli came first. But, as has already been pointed out, the question of which came first could not even have been asked in cases where the successive flashes of light stimulated the same place on the retina or where two Savart-wheel clicks followed each other. Those cases that Fraisse (1957) regards as difficult for obtaining judgments of simultaneity are the very ones in which the question of temporal order may be asked because the successive stimuli are different in quality, in place, or in sense modality. One of the conclusions of the present paper will be that the important distinction here is not between pairs of stimuli delivered to one sense modality as opposed to pairs delivered to separate sense modalities, as Fraisse (1957) and Piéron (1952) have suggested, but rather the more basic distinction is between experiments that ask about successiveness as opposed to simultaneity and those that require the judgment of temporal order. This conclusion will be justified if one can return to conditions of stimulation in one sense modality alone with differentiable stimuli whose order must be judged correctly and in which the results yield numbers more like those derived from heterogeneous stimulation where one has little choice but to ask for judgments of order.

Auditory perception of temporal order.—If two stimuli are presented to the same sense modality, are identical in quality and intensity, and occur in the same place, then one cannot ask about order of occurrence because there is no way for *O* to label separately each of the two stimuli. Here one is limited to asking whether there appear to be one or two. In a previous experiment in hearing (Hirsh, 1959), various pairs of sounds were manipulated with respect to time of occurrence, and in each pair the individual sounds were made different from each other with respect to frequency (if they were tones), spectrum (if they were bands of noise), and duration (if the pair consisted of both a brief and a prolonged sound). Under these circumstances, we could ask *O* which sound came first—the higher or lower pitch, the longer or shorter sound, the tone or noise, etc. The surprising result from these various combinations of sounds was that, whereas a 2-msec. separation is sufficient to yield nonsimultaneity, the time intervening between the two sounds that was required for judgments that were 75% correct was of the order of 20 msec. and was independent of the kinds of sounds used. The generality of this finding within the auditory domain suggested the present series of experiments whose aim it was to extend these observations to other sense modalities and also to pairs of stimuli that involved more than one sense modality.

EXPERIMENT I: VISUAL ORDER

The question for Exp. I was: how much time must intervene between two visual stimuli for *O* to report correctly the order in which they occur? The two stimuli were identical except that they appeared at different places—at two points along either a horizontal or a vertical line. The *O* was asked to report on whether the left or right or the upper or lower light came first; he could not say "simultaneous."

Procedure

Flashes of light were generated by electrical pulses of 5 msec. duration that fed an amplifier circuit in the output of which was connected a Sylvania R-1131C glow modulator tube. This glow tube was mounted in a box, its end facing *O*, behind a ground Plexiglas screen. A homogeneous illumination was achieved over a circular area of 0.5 cm. diameter. Level of illumination was set at 30 db. above absolute threshold. The pulses were generated by pulse generators (Tektronix Model 160) which in turn were driven by a single waveform generator (Tektronix Model 161) in such a way that the time interval between the two pulses could be manipulated with an accuracy of ±1 μsec. The sequence of two pulses was repeated once per second.

The visual targets were placed on a table and, for the horizontal displays, were spaced 18, 36, or 72 cm. apart. With *O* 2 m. away, these separations corresponded to visual angles of 5°, 10°, and 20°. When the visual targets were disposed vertically, they were separated by 18 cm. The three horizontal distances were used with a red fixation point in the center so that the two white lights whose order was to be judged were to the left and to the right of center. In addition, an 18-cm. horizontal separation was used with the fixation point 18 cm. to the left of the left target and also 18 cm. to the right of the right target. The 18-cm. vertical separation was used with the fixation point both 18 cm. to the left of the vertical line and also 18 cm. to the right. These various conditions were employed to investigate the dependence of judgments of visual order upon the region of the visual field stimulated and also the spatial direction of the change.

One particular spatial arrangement was used in each experimental session. During each session 10 presentations of 7 different

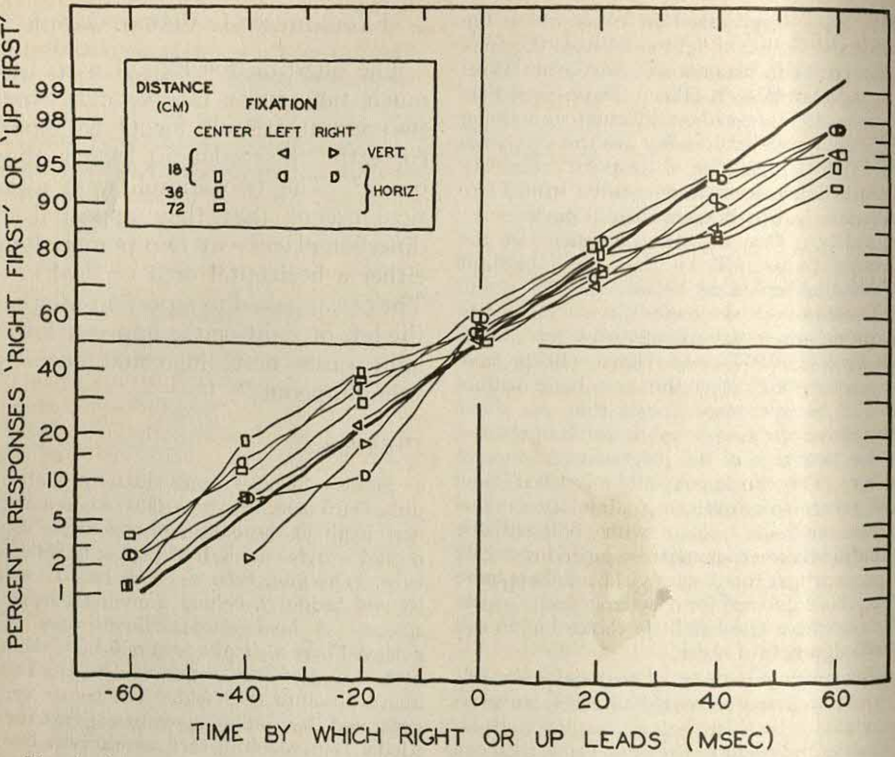


FIG. 1. Visual temporal order. Average percentage of responses "right first" or "upper first" as a function of the time by which the right or upper light preceded the left or lower light in milliseconds. (Parameters, indicated by different symbols as defined in the key, represent different spatial arrangements of the two lights. The heavy straight line comes from data on auditory perception of temporal order [Hirsh, 1959].)

temporal intervals were randomly ordered. The intervals ranged from 60 msec. with the left or lower light leading, to 60 msec. with the right or upper light leading, in 20-msec. steps. Four trained *O*s, between 18 and 30 yr. old, were tested singly. All sessions were repeated a second time with a different order of conditions for each *O* so that the results are based upon 20 presentations of each of the 7 temporal intervals under each of the 7 spatial arrangements.

Results

The average percentage of responses "right first" for horizontal displays and "upper first" for vertical displays is plotted in Fig. 1 as a function of the time by which the right or upper light leads the left or lower light. (Negative values on the abscissa thus represent the time by which the right or upper light lagged.) The different symbols denote the different distances and spatial arrangements of the visual targets. On the basis of a graphic inspection alone it appears that there are not systematic differences related to different spatial arrangements. The heavy line in Fig. 1 is the line that represents judgments of temporal order for auditory stimuli, being a summary of the several combinations of sounds used in a previous study (Hirsh, 1959). This function, which describes the auditory analog of this experiment, may be slightly steeper than a line that would be drawn through the average of the present visual data, but there is not a great difference between the two, certainly not as much difference as one would anticipate on the basis of the relative speeds with which visual and auditory systems are supposed to act.

Results for mere successiveness (as opposed to fusion or simultaneity) reveal larger separation times for vision than for hearing. When *O* is asked to report on the order in which successive events occur (rather than on mere successiveness), however, the differences between these two sense modalities are small, if indeed they are real.

EXPERIMENT II: AUDITORY ORDER AND SPACE

Experiment I differed from the previous studies on hearing (Hirsh, 1959) in that in the auditory experiment the two events whose order was to be judged differed from each other with respect to qualities like pitch, duration, etc., while in Exp. I, the two events whose order were to be judged were qualitatively the same visual stimulus but occurred at two different places. In spite of the relatively good agreement between the visual and auditory results, we were interested to know whether the auditory results would obtain if the two auditory stimuli whose order was to be judged were qualitatively the same but occurred at two different places.

Procedure

Pairs of pulses of 0.1 msec. duration were produced by the same system of pulse and timing generators that were used in Exp. I. These pulses were led separately to electrical simple tuned circuits. The resonant frequency of these tuned circuits was varied through the use of different fixed capacitors and the damping was also varied through the use of a variable resistor.

Three different conditions of stimulation were employed in order to make the two sounds whose order was to be judged (*a*) different with respect to pitch, (*b*) different with respect to which ear was stimulated, and (*c*) a combination of the first two.

The first of these conditions constitutes a repetition of one of the conditions in the previous auditory work (Hirsh, 1959). Here both pulses were presented to the same ear. One was passed through a circuit resonant at 666 cps and the other through a circuit resonant at 278 cps. The damping in both circuits was so adjusted that the amplitude of the damped oscillation that characterized the now-filtered pulse decreased to $1/e$ of its original value in 10 msec. Both pulses now sounded like brief xylophone notes of high and low pitch, which will be referred to hereafter as the "bing" and the "bong" respectively. Five new *O*s, trained young adults, were tested singly and asked to report which of the two pitches came first, the higher or the lower. Bings and bongs were separated by 10, 20, and 30 msec., with the bing leading one-half of the time and the bong leading the other half. A random series of 60 trials, 10 presentations at each of the six temporal separations, comprised an experimental session. No stimuli were presented with 0-msec. separation because a characteristic quality of simultaneity was always observed.

The second of the three conditions was intended to duplicate for the auditory system the spatial separation that characterized the visual experiment. In this case the two pulses were passed through the same high-frequency circuit (666 cps), one being led to the left ear and the other to the right ear. (We presume that this is Exner's [1875] two-ear condition.) The stimuli were now qualitatively the same, the difference between them being only the side on which the sound was heard. The same six temporal intervals were used. No stimuli were delivered with 0-msec. separation because we did not wish to confound the judgment of perceived order with the intracranial localization of a fused image. It should be clear that the temporal intervals used were all much longer than are involved in localization within the head or even fusion of the two stimuli into one sensation. Again a random series of 60 presentations, 10 at each of the six temporal separations, comprised an experimental session. Each of the 5 *O*s was asked to report which sound came first, the one on the right or the one on the left.

Finally, the two dimensions in which the auditory stimuli were different in the first two conditions were combined so that a high-frequency bing was presented to the right or left ear while the low-frequency bong was presented at the left or right ear, and *O* was asked to report which came first, the sound on the right or the sound on the left. His response was still based upon location in space, even though the two events that occurred at the different places were also qualitatively different with respect to pitch.

Results

The results of these three subexperiments are shown in Fig. 2. The open circles show the percentage of judgments "higher first" as a function of the time by which the high-pitched bing leads the lower-pitched bong. The line connecting these open circles points to a finding that was reported on previously, namely that brief stimuli may in fact generate a function that is somewhat steeper (i.e., yields more precision) than the average for all auditory pairs. Judgments concerning which of two qualitatively identical bings comes first when they are delivered separately to the two ears are shown by the Xs. These points do not appear to deviate greatly from the straight line which again represents an average of results on perceived order for all sound pairs used in Hirsh's (1959) previous study. The closed circles show similar results for the condition in which the two events whose order is to be judged are different with respect to both place and quality. The lines connecting these points indicate a relatively broad range of indeterminacy around 50% between temporal separations of ±10 msec. Once beyond this range, however, the rate at which judgments change with increasing temporal separation appears to be greater than that for other conditions.

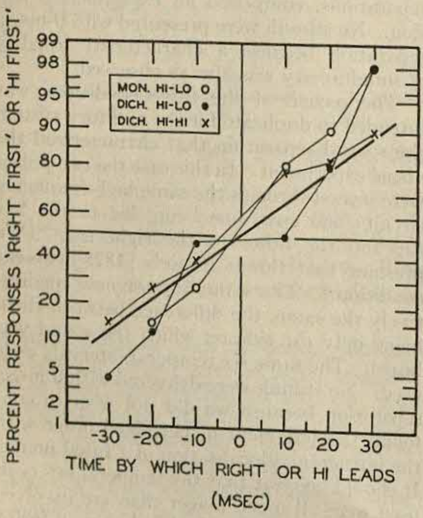


FIG. 2. Auditory temporal order. Average percentage of responses "right first" or "higher first" as a function of the time by which the right ear or the higher pitch preceded the left ear or lower pitch in milliseconds. (The open circles show results for monaural perception of order between two different pitches; the Xs show results for perception of order between the ears receiving the same pitch, and the filled circles show results for perception of order between two events that are different both with respect to ear and pitch. The straight line summarizes previous results in auditory perception [Hirsh, 1959].)

EXPERIMENT III: PERCEIVED ORDER AT TWO PLACES

So far we have demonstrated, among other things, that approximately the same amount of time is required to tell which of two stimuli came first when those two stimuli are placed on different sides, either in external space (visual) or in the head (auditory), no matter whether the visual or auditory modality is stimulated. To test the generality of this conclusion, we set up a similar experiment for the tactual system, using vibratory stimuli to the index fingertip of the right and left hands. The 5 *O*s were the same as those used in Exp. II but different from those used in Exp. I. We proposed to compare the percentage of responses "right first" as a function of the time by which the right led the left in three sense modalities: auditory, visual, and tactual. The first third of the experiment was already done and comprises the data on right versus left clicks in Exp. II (the X symbols in Fig. 2).

Procedure

The apparatus and procedure used for the visual part is exactly the same as discussed under Exp. I. Only one distance was used, 18 cm., the two target lights being on the same horizontal line with a red fixation point halfway between. In the same group of 5 *O*'s, each *O* was asked to say whether the right or left light came first as the temporal separation between the two lights was varied from 30 msec., with the left leading, to 30 msec. with the right leading, in 10-msec. steps.

The outputs of the same pulse generators were fed to a power amplifier whose output was delivered to two Goodmans V-47 vibrators. These somewhat resonant transducers damped the 1-msec. rectangular pulse in such a way that the mechanical output, when the transducer was loaded by a fingertip, resembled approximately a half-wave at 100 cps. The *O* sat with his arms resting on two tablet-arm chairs and his index fingertips resting lightly on the vibrators. His wrists could remain on the tablet so that only the weight of his finger was supported by the vibrator. His task was similar to that in the other experiments: pairs of stimuli were delivered to the two vibrators and he was asked which of the two fingers was stimulated first, the right or the left. Temporal intervals between the tactual stimuli were varied from 30 msec., with the left leading, to 30 msec. with the right leading, in 10-msec. steps.

Results

Several attempts were made in preliminary experiments to collect data on the temporal resolving power of the tactual system. The *O*s were asked to say which of the following paired locations were stimulated first: face or toe, finger or palm, finger or arm. Wide variations among *O*s were observed, partly because this particular range of temporal intervals serves the tactual sense in much the same way as temporal intervals of the order of microseconds serve auditory localization. For example, when the fingertip and the palm are stimulated successively, a time interval of from 20 to 40 msec. appears to be sufficient for *O* to say which spot was stimulated first, but for a temporal interval only somewhat smaller, for example about 15 or 20 msec., there is but one stimulus perceived and it is located neither at the fingertip nor at the palm, but at some place in between, depending upon the interval. Data from these preliminary studies are at best equivocal because of the wide variations among *O*s and will not be reported here.

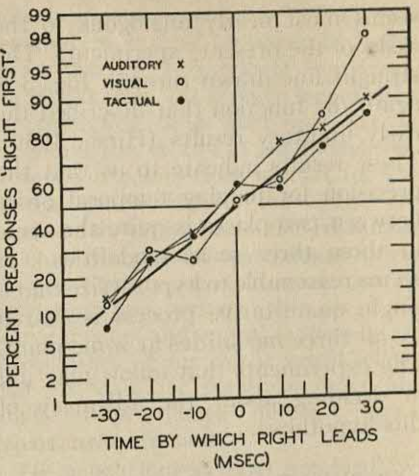


FIG. 3. Temporal order in space. Average percentage of responses "right first" as a function of the time by which the right stimulus precedes the left stimulus in milliseconds. (The different symbols represent three different sense modalities. The straight line summarizes previous results in auditory perception [Hirsh, 1959].)

The particular task described above is not so difficult if the places for tactual stimulation are on opposite sides of the body, namely the fingertips. Results averaged over the 5 *O*s are shown by the filled circles in Fig. 3. The visual experiment in which the targets were placed 9 cm. to the left or to the right of a fixation point yielded results shown by the open circles in Fig. 3. The Xs are taken directly from Fig. 2 and are used for comparison with tactual and visual data because that particular auditory task (which ear was stimulated first by the same quality click) seems most nearly analogous to the tasks of the present experiment. The straight line drawn through Fig. 3 is again the function that described the early auditory results (Hirsh, 1959). These results indicate to us that the precision for judging temporal order between two places is quite the same in these three sense modalities. It seems reasonable to hypothesize that a single quantitative process is shared by all three modalities at some point. The experiments that follow may aid in deciding about the adequacy of this hypothesis.

EXPERIMENT IV: PERCEIVED ORDER BETWEEN MODALITIES

One further experiment seemed in order before we could establish the generality of data on perceived order and their independence of particular sense modalities. It has been noted that the classical literature on temporal resolving power did indeed treat the problem of perceived order of stimulation in two sense modalities (Exner, 1875). These data have led some to hypothesize that sensations in two sense departments cannot be so well organized in time as sensations in the same sense. But the data for the better temporal organization in single sense modalities have come from judgments of successiveness (as opposed to simultaneity) rather than from judgments of order. If our notion is correct, then the results indicate merely how much more time is required for judgments of perceived order when stimuli are presented to the same sense modality. Now we wished to test whether similar behavior would be observed when the two stimuli in each pair were delivered to different sense modalities.

Procedure

The apparatus has already been described. The additional variation here was that the output of one pulse generator was transduced by a glow modulator for a visual stimulus, while the second pulse was transduced by the vibrator for a tactual stimulus. By varying the temporal interval between the two pulses, we could vary the time by which the visual flash preceded or followed the tactual vibration. In this case, *O* was asked which came first, the tap or the flash.

When either the visual or tactual stimulus was to be paired with the auditory, then one of the pulses was filtered (666-cps bing) and transduced by an earphone.

The same 5 *O*s of Exp. II and III were tested singly. One session was devoted to a particular pair of modalities: either auditory-visual, visual-tactual, or auditory-tactual. Under each of these conditions, 70 stimuli were used, 10 presentations at each of 7 temporal intervals. All stimuli were maintained at a level 30 db. above threshold.

Results

Results are shown in Fig. 4. The different symbols refer to the three pairs of modalities. The open circles

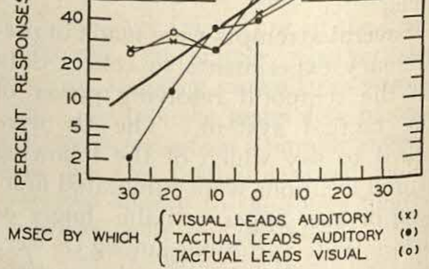


FIG. 4. Perceived order of stimuli affecting different sense modalities. (The ordinate shows the average percentage of responses "visual first" as a function of the time separating auditory and visual stimuli [Xs]; or "touch first" as a function of the time separating auditory and tactual stimuli [filled circles]; or "touch first" as a function of the time separating visual and tactual stimuli [open circles]. The straight line summarizes previous results in auditory perception [Hirsh, 1959].)

show the percentage of responses that were "touch first" as a function of the time by which the tactual stimulus led the visual stimulus. The negative numbers on the left side of the abscissa then would, for this particular set of symbols, indicate that the visual is leading the tactual stimulus. Similarly, the closed circles show the percentage of responses that were "touch first" as a function of the time by which the tactual led the auditory stimulus. Finally, the Xs show the percentage of responses that were "visual first" as a function of the time by which the visual flash led the auditory click. The straight line is again a summary of Hirsh's (1959) previous auditory experiments, and except for a few departures at the lower left, the points for the present three subexperiments seem not to deviate greatly from the same function.

DISCUSSION

Differences among the auditory, visual, and tactual modalities with respect to temporal resolving power are well known. Judgments of successiveness as opposed to simultaneity, in the case of discrete stimuli (the temporal 2-point limen), or in the case of a long series of discrete stimuli (fusion), show that the visual system is relatively sluggish while the tactual system is less so, and the auditory system seems to be the fastest.

More than such simple resolving power is required when *O* is asked which of two sensations precedes the other. Temporal resolution for successiveness is, to be sure, a necessary prerequisite for this judgment of order, but it is not sufficient. The judgment of order requires that two pieces of information must be organized with respect to time, and the results of the present studies seem to indicate that it does not much matter where these two pieces of information come from; that is, they may come from different parts of the same sensory mechanism or they may even come from different sense modalities. The time required to insure, for example, judgments that are 75% correct is approximately 20 msec. It is much longer than the resolution times that give rise to successiveness in any of the modalities, although it comes close to that for the slowest system, namely the visual. It is so much longer than the resolution times of the several sense departments that it seems to point to some kind of time-organizing system that is both independent of and central to the sensory mechanisms. From this it would follow that the data for perceived order are similar even though different sense departments are used.

The much longer times obtained by Exner (1875) in his experiments can be accounted for partly by the psychophysical method he employed. At one point Exner described his method as one in which a judgment of correct order *10 times successively* was required to fix the value of time for perceived order. From the present data, such a stringent requirement would place the time values for all conditions at greater than 40 msec. This would be about the 98% point on the psychophysical functions in the figures shown. A word should be said about *O*'s task in the experiments. When *O* is asked to say that one perception equals another, or as in this case coincides with another, it is commonly found that he will show some variation in preference for the response "equal" over responses "greater," "less," etc. In the present experiments, it was found that *O*s in training complained repeatedly that the perceptions were simultaneous. They were forced to decide on one order or the other, however, and the result was as shown. This is not a new fact in psychophysics, but it has not been clearly stated heretofore in connection with the data on temporal order. Had the present writers chosen the method of average error as the technique for obtaining thresholds, it is undoubtedly true that the *SD* would have been considerably greater than 20 msec.

One of the exceptions to the function

that we have come to regard as a general rule, namely the straight line whose intercept at 75% is 20 msec., is the case of two brief auditory stimuli whose order is to be judged. In the previous auditory study (Hirsh, 1959), there was some evidence that the slopes of analogous functions, where clicks were paired with tones, were somewhat steeper than the average slope which was more typical of judged order for the onsets of tones and noises. This is also seen in the bing-bong auditory study represented in Fig. 2. Broadbent and Ladefoged (1959) have suggested that Hirsh's (1959) original results which indicated that perceived order could be judged correctly with as little as 20 msec. intervening between sounds, were a combined result of highly trained *O*s and the fact that when the temporal interval between two sounds that are already close together in time is changed, the overall quality of the two-sound complex is changed. They suggested that it was on this basis that *O*s were making their judgments and this may, in fact, be the case, especially for brief stimuli where the slope of the function is even steeper. However, this kind of judgment of complex quality cannot be the case in the further experiments presented here. To be sure, the time intervals used in the visual experiment lay in a range where sometimes apparent movement was seen and there the judgment of "which came first" may have been deduced from "in what direction did the light move." When the different sense modalities were stimulated, however, there is no impression of movement; there is rather only a first and a second—and the same function seems to fit the data equally well.

SUMMARY

A series of studies has been carried out in which it has been shown that the amount of time that must intervene between two events in order for *O* to report correctly (75% of the time) which of those two events preceded the other is approximately 20 msec. The function that relates percentage of judgments to the temporal interval between the events appears not to be different even though the two events are either two sounds of different quality, or two light flashes in different places, or two vibrations, one to each hand, or two sounds, one to each ear, or two different stimuli, one to each of two sense modalities.

Whereas the time between successive stimuli that is necessary for the stimuli to be perceived as successive rather than simultaneous may depend upon the particular sense modality employed, the temporal separation that is required for the judgment of perceived temporal order is much longer and is independent of the sense modality employed.

REFERENCES

BARTLEY, S. H. *Principles of perception.* New York: Harper, 1958

BROADBENT, D. E., & LADEFOGED, P. Auditory perception of temporal order. *J. Acoust. Soc. Amer.*, 1959, **31**, 1539.

EXNER, S. Experimentelle Untersuchnung der einfachsten psychischen Processe. *Pflüg. Arch. ges. Physiol.*, 1875, **11**, 403–432.

FRAISSE, P. *Psychologie du temps.* Paris: Press Univer. France, 1957.

GELDARD, F. A. Perception of mechanical vibration. *J. gen. Psychol.*, 1940, **22**, 243–269.

HIRSH, I. J. Certain temporal factors in audition. *Science*, 1952, **116**, 523.

HIRSH, I. J. Auditory perception of temporal order. *J. Acoust. Soc. Amer.*, 1959, **31**, 759–767.

JAMES, W. *Principles of psychology.* Vol. 1. New York: Holt, 1890.

LADD, G. T. *Outlines of physiological psychology.* New York: Scribner, 1892.

MILLER, G. A., & TAYLOR, W. G. Perception of repeated bursts of noise. *J. Acoust. Soc. Amer.*, 1948, **20**, 171–182.

PIÉRON, H. *The sensations.* (Trans. by Pirenne and Abbott) London: J. Garnet Miller, 1952.

TROLAND, L. T. *Principles of psychophysiology.* Vol. 1. New York: Van Nostrand, 1929.

WALLACH, H., NEWMAN, E. B., & ROSENZWEIG, M. R. Precedence effect in sound localization. *Amer. J. Psychol.*, 1949, **62**, 315–336.

(Received September 1, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 433–438

THE PREDICTION OF FREE RECALL FROM WORD ASSOCIATION MEASURES

ERNST Z. ROTHKOPF AND ESTHER U. COKE

Bell Telephone Laboratories

This is a report of some observations on frequency and sequence effects in the free recall of individual items from a word list and on the relationship of these effects to data from word association norms. Such observations are not new in the literature. Sequence effects ("clustering") have been described by Jenkins and Russell (1952) and by Jenkins, Mink, and Russell (1958). Frequency of recall effects have been reported by Deese (1959a, 1959b) who showed that nonlist words which are common association responses for the words of a list tend to occur as extralist intrusions and that this effect is an increasing function of the average association frequencies between the words which compose the list and the intruding term. Deese (1959a, 1960) has also shown that more words were recalled from lists with high degree of intralist associations than from lists of low degree of intralist associations.

The studies referred to above have concerned themselves with overall list characteristics and not with the recall of the various individual words within a list although the first Deese study (1959b) indirectly bears on this problem. The prediction of recall, from word association norms, for the various individual words within a list is somewhat complicated by the fact that interassociations among list words are generally complex and that no satisfactory rational basis for weighting these interassociations is now available. The present experiment was intended as a test of a reasonably simple scheme for predicting the recall of individual words from word association norms. This scheme involves the use of the number of words in the list to which a given word, W, is normally given as an association response, as a predictor for the frequency of recall of W. It grew from the hypothesis (Bousfield, 1953) that free recall was in part dependent on cuing ("relatedness increment") among list components. The scheme was tested on a list composed of 99 words from the Kent-Rosanoff Word Association Test.

METHOD

Materials.—The list of stimulus words which were used in the free recall task consisted of 99 words from the 100-word Kent-Rosanoff list. One of the words, *cheese*, was omitted because procedural convenience required that the number of words in the lists be divisible by three. The Kent-Rosanoff words were chosen because recent word association norms based on large numbers of *S*s were available (Russell & Jenkins, 1954). Each stimulus word was typed in capitals on a separate 3 × 5 in. card.

Procedure and subjects.—Three *S*s were tested at the same time. The deck of stimulus cards was shuffled and split into three numerically equal subdecks. One of these was placed face down before each *S*. At *E*'s signals, *S*s inspected each card of their subdeck for 5 sec. When *S*s had finished viewing their subdeck, *E* reshuffled each subdeck, reassigned them, and inspection was continued. This procedure was repeated until each *S* had inspected each of the 99 words, once, for 5 sec.

Immediately following a 15-min. rest pause, *S*s were asked to write down all the words of the stimulus set which they could remember. A period of 5 min. was allowed for recall. The *S*s wrote the recalled words on 3 × 5 in. paper pads, one word to each page.

All *S*s ($N = 28$) were volunteer eleventh and twelfth grade students who were paid for their services. All *S*s were taking a college preparatory course. They had participated on the previous day in an identification learning experiment which did not involve words either as stimuli or as response terms.

Results

The number of words in the list for which a given word, W, is normally given as an association response will be referred to as W's *cue number* or N_C. N_C was determined by counting all stimulus words (except "cheese") for which W was given as an association response in the 1952 Minnesota norms of the Kent-Rosanoff Word Association Test (Russell & Jenkins, 1954). Only those association responses were considered which occurred either with a normative frequency of more than 11 or which were among the 10 most common association responses to the stimulus word, whichever yielded the greater number of different association responses. N_C was obtained separately for each word of the list. The frequency distribution for words of various N_C in the list is shown in Table 1 (for additional intralist association data for the 99 word K-R list see Rothkopf & Coke, 1961). The incidence of words for which $N_C = 0$ is 26. Frequency of incidence in general decreases as cue number (N_C) goes up.

TABLE 1

INCIDENCE OF WORDS WITH VARIOUS CUE NUMBER (N_C) IN THE KENT-ROSANOFF LIST (MINNESOTA 1952 NORMS)[a]

N_C	Frequency	N_C	Frequency
0	26	9	1
1	23	10	2
2	18	11	3
3	11	12	1
4	3	13	0
5	2	14	0
6	1	15	1
7	3	16	1
8	3	>16	0

[a] The word "cheese" has been omitted.

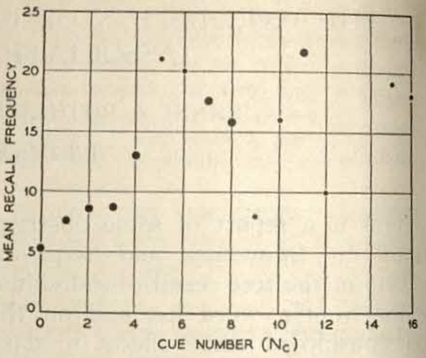


FIG. 1. Mean recall frequencies for words of various cue number. (The smaller datum points involve only three observations or less.)

Mean frequency of recall for words of the same N_C value was plotted against N_C in Fig. 1. As indicated before fewer observations were available for the high than for the low N_C values. The high N_C portions of the Fig. 1 plot may therefore be less reliable estimates of the frequency of recall function than the low N_C portions. For this reason homogeneity of variance throughout the N_C range was tested by the Bartlett technique prior to proceeding with an analysis of variance. The five N_C values for which only one observation was available were omitted from this analysis. The results of this test did not permit the rejection of the homogeneity hypothesis ($P > .3$). The analysis of variance, summarized in Table 2, shows that variance due to N_C was significantly greater than within-N_C variance ($F = 7.01$, $P < .01$). The linear component of between-N_C variance was significant ($F = 70.23$, $P < .01$). However, the remainder variance was also found to be significantly larger than the within variance although only at the .05 level ($F = 2.14$). An inspection of the plot suggests that there may be some negative acceleration of the function. It seems reasonable to argue that there is a monotonic, increasing relationship between recall and N_C. The product-moment correlation between recall frequency and N_C was .62. A simple attempt to weight N_C by association frequencies did not improve the correlation with recall. Weighting of N_C was accomplished by summing the association frequencies with which W is given for its various stimulus words. These sums when correlated with recall yield an r of .54. The association frequency sums are, of course, highly correlated with N_C ($r = .84$).

In view of the fact that recall of words and their frequency of use are positively related (see Underwood & Schulz, 1960, pp. 56–57), a serious possibility exists that the function shown in Fig. 1 was the result of a correlation between N_C and frequency of use. In order to check this possibility correlations were computed between N_C, frequency of recall, and frequency of use as estimated from the Lorge-Thorndike semantic count (Thorndike & Lorge, 1944). The obtained r between N_C and use frequency was .41; r between use frequency and recall was .18. The partial correlation between frequency of recall and N_C with the effect of frequency of use held constant was .609. It must be concluded that while N_C and frequency of use are to some extent related, the effect of this relationship on the correlation between recall and N_C is very small.

TABLE 2

Analysis of Variance of Frequency of Recall of Kent-Rosanoff Words as a Function of Cue Number (N_C)

Source	*df*	*MS*	*F*
Between N_C	14	142.16	7.01*
linear	1	1424.27	70.23**
residual	13	43.54	2.14*
Within N_C	84	20.28	
Total	98		

* $P < .05$.
** $P < .01$.

If the effect of N_C on recall is to be accounted for in terms of a cuing mechanism, it should be expected that words will be recalled immediately after any one of the words for which they are normally given as association responses with frequencies which are in excess of chance. It should also be expected that the incidence of such sequences will increase with N_C even when the fact that words with high N_C are recalled more often than words with low N_C is taken into consideration. This was tested in the following manner. Whenever a given word followed, in free recall, immediately after a word for which it was an association response, this event was referred to as a *cue linkage*. In order to adjust for differences in recall frequency among words, the frequency of cue linkages in the recall protocols of all *S*s for every word was divided by the number of opportunities for cue linkage for that word. Opportunities for cue linkage were calculated by subtracting the number of times a word was recalled first (when no linkage could occur) from the total frequency with which that word was recalled. The mean ratio of cue linkage to opportunities for cue linkage is plotted against N_C in Fig. 2. The ratio of cue linkage to opportunities for cue linkage is an increasing function of N_C. This conclusion was supported by an analysis of variance, summarized in Table 3. Prior to performance of the analysis of variance, Bartlett's technique was used to test the hypothesis of homogeneity of variance throughout the N_C range. Again all

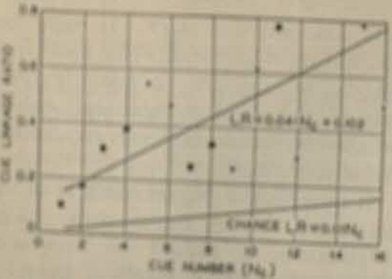


Fig. 2. Cue linkage ratio (cue linkage/opportunity for linkage) as a function of cue number. (The top line has been fitted to the data by the method of least squares. The lower line—chance function—is explained in the text.)

N_c values involving only one observation were not included. The homogeneity of variance hypothesis could not be rejected ($P > .9$). Table 3 shows a significant F ratio for between-N_c variance ($F = 5.61$, $P < .01$), and also indicates that the linear component of between-N_c variance was significant ($F = 50.24$, $P < .01$). The slope of the empirical function in Fig. 2 is significantly greater ($t = 5.037$, $df = 71$, $P < .01$) than the slope of the chance function which is also shown in Fig. 2. The chance function was based on the fact that words with greater N_c are more likely, by chance, in terms of the scoring criterion, to follow words for which they are association responses. The chance function was calculated by Chance Function = $N_c/98$. The finding that cue linkage increases with N_c even when differences in recall frequencies and chance factors are compensated for, is consistent with the hypothesis that the positive relationship between N_c and recall depends on some cuing mechanism.

TABLE 3

Analysis of Variance of Cue Linkage: Opportunity for Linkage Ratio as a Function of Cue Number (N_c)

Source	df	MS	F
Between N_c	13	.1811	5.61**
linear	1	1.6226	50.24**
residual	12	.0611	1.89
Within N_c	59	.0323	
Total	72a		

a There are 26 fewer df here than in Table 2 because Table 3 analysis did not include 26 words for which N_c was zero since for these words no cue linkage can occur.
** $P < .01$.

The following procedure was used to find whether cue linkage depended on association strength (frequency) between the cuing word and the link (cued) word. The association norms for each cue word were examined and all association responses which were stimulus words were noted and their frequencies recorded. It was determined for each cue linkage whether the association frequency of the linked word was above, equal to, or below the median association frequency of the relevant associations to the cue word. There were 265 cue linkages in 1000 recalled words. Of these 265, 178 were above median, 30 were equal to the median, and 57 were below the median. The 30 equal-to-the-median cases were equally divided between the above and the below median category and a binomial test was performed (Siegel, 1956, pp. 36-41). The two-tailed test was significant ($z = 7.31$, $P < .01$).

An estimate of the amount of organization which was imposed on sequence of recall through the associations among the stimulus words was obtained by comparing cue linkage scores for individual *Ss* (hence referred to as AA, actually recalled words in actual order), with two different estimates of chance cue linkage scores. One chance cue linkage score, which will be referred to as AR (actually recalled words in random order), was obtained by taking the words actually recalled by each *S*, arranging them in random order, and counting the cue linkages which occurred. Ten random arrangements were made for each *S*'s set of words. The mean cue linkage over the 10 random arrangements was used in the comparison tests. The second chance linkage estimate, referred to as RR (randomly chosen words in random order) was generated by randomly choosing words from the 99 word list. The number of words chosen was the number of words actually recalled by an *S* less the number of nonlist responses made by that *S*. That number of "nonlist" terms[1] was then added to the randomly chosen words and the resulting set was then arranged in random order. This process was repeated 10 times for each *S*'s actual recall frequency and the mean of these 10 values was used in the statistical analysis. Mean cue linkage score for the actual recall protocols (AA) was 9.50. Mean linkage for the AR and RR were 1.84 and .84, respectively. Because of marked heterogeneity of variance, Friedman's analysis of variance by ranks (Siegel, 1956, pp. 166-172) was used to compare the three conditions. The overall comparison as well as the AA:AR and AR:RR differences were highly significant ($P < .001$).

DISCUSSION

The results of this experiment support the hypothesis that the frequency of recall of individual words within a list is at least in part determined by the number of list words (N_C) for which the word in question is normally given as an association response. This finding is consistent with previous observations reported by Deese (1959a) which linked average frequency of recall for lists with interassociations among list items.

Another finding of this experiment was that the frequency with which a word is recalled immediately after one of its "cues" is an increasing function of the number of "cues" for the word in question. This result, which held even when differences in recall frequency and chance factors were taken into account is consonant with earlier findings by Jenkins et al. (1958; Jenkins & Russell, 1952) that associates tend to occur consecutively in recall.

The comparisons among cue linkages in actual recall sequences (AA) and the computer generated chance linkage estimates (AR and RR) support the conclusions which were presented in the preceding paragraphs. The finding that AA data show more linkages than the generated AR sequences indicates that *S*s tend to recall associates in immediate sequence more often than can be expected by chance. The fact that the generated AR sequences result in greater cue linkage frequencies than the RR sequences implies that words which are actually recalled by *S*s are potentially more linkable, i.e., have higher N_C values, than words which are randomly chosen from the list. This conclusion follows from the data because both AR and RR involve random sequences which differ only in that the AR sequences are composed of actually recalled words while RR sequences were generated by random choices from the 99 word list.

Taken together, the results of this experiment are consistent with a single, cuing interpretation of both the frequency of recall and the sequencing phenomena which were the subject of this study. Such an interpretation has been proposed by Bousfield (1953) who calls the cuing effect "relatedness increment." A reasonable model for free recall which, when elaborated, may fit the present data and also the findings of Jenkins et al. (1958; Jenkins & Russell, 1952) and Deese (1959a, 1959b, 1960) may include the following elements. At any time in recall that word among the not yet recalled words which has the greatest superthreshold habit strength is recalled next. When a word is recalled it increases the habit strength of all its association responses by amounts which are proportional to the association frequencies. In this way the position of the cued words in the hierarchy of habit strengths of words in the list may be temporarily changed. Cuing may also bring below-threshold habits temporarily to above-threshold strength. The failure to improve the prediction of free recall from cue number by weighting N_C by association frequency may be due to the fact that the words of the Kent-Rosanoff list differ in many other respects besides interassociations. It is also possible and perhaps likely that available word association norms do not provide adequate indices of interitem association strength and that new indicators need to be sought.

---

[1] The mean number of nonlist words in recall was three per *S*. The number of nonlist responses did not vary appreciably with amount of recall.

SUMMARY

It was hypothesized that the frequency of recall of a given word W in a free recall situation depends directly on the number of cues which are provided by other words in the list. The cue number of word W was defined as the number of words in the learning list for which W is normally given as an association response in a word association test. The hypothesis was tested by exposing *S*s ($N = 28$) to each of 99 words of the Kent-Rosanoff list for 5 sec. and testing for recall after 15 min.

The data supported the following conclusions: (*a*) The frequency with which a word was recalled was an increasing function of its cue number. (*b*) The frequency with which a word was recalled immediately after one of the words for which it was an association response (cue linkage) was an increasing function of its cue number even when the effects of recall frequency and chance differences due to cue number were taken into account. (*c*) Words which followed one of their cues in recall tended to be among those word associations which had association frequencies above the median association frequency of the association responses (list words only) to the cue word. These results were consistent with a cuing interpretation of the frequency and sequence of recall phenomena which were the subject of the present study.

REFERENCES

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312. (a)

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22. (b)

DEESE, J. Frequency of usage and number of words in free recall: The role of association. *Psychol. Rep.*, 1960, **7**, 337–344.

JENKINS, J. J., MINK, W. D., & RUSSELL, W. A. Associative clustering as a function of verbal association strength. *Psychol. Rep.*, 1958, **4**, 127–136.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–821.

ROTHKOPF, E. Z., & COKE, E. U. Intralist association data for 99 words of the Kent-Rosanoff word list. *Psychol. Rep.*, 1961, **8**, 463–474.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. N8-ONR-66216)

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

UNDERWOOD, B. T., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Early publication received March 27, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 439–447

CONDITIONAL RESPONSE PROBABILITY IN A T MAZE[1]

ROBERT S. WITTE

San Jose State College

In a previous study (Witte, 1959), a trace-fluctuation model derived from statistical learning theory was applied to choice behavior under sequentially dependent reward schedules. An adequate test of the theory was hampered by the large number of *S*s with position preferences. The present study attempts to test the theory further by using a procedure designed to yield within-*S* comparisons, as well as to minimize position preferences.

When reinforcement is independent of *S*'s responses and occurs on every trial, the trace-fluctuation model generates the well-known "probability-matching" prediction for unconditional response probability. When, in addition, trials are massed, the model predicts that conditional response probability will be a direct function of the sequential dependencies between reinforcements. More specifically, the asymptotic probability of a response similar to the one reinforced on the previous trial, henceforth called a "repetition" response, following Anderson's (1960) usage, is

$$p_{ii}(\infty)=\frac{1}{N_c+N(1-j)^t}\{N_c[\pi+\theta(1-\pi)] + N(1-j)^t[\pi_{ii}+\theta\pi_{ii}(1-\pi_{ii})]\} \quad [1]$$

where $p_{ii}(\infty)$ is the asymptotic probability of a repetition response; N_c is the number of elements in the set of common stimulus elements; N is the number of "trace" elements in the set associated with an E_i reinforcing event; j is the rate at which the N trace elements become inactive; t is the size of the time interval since E_i occurred; π is the unconditional probability of E_i; θ is a sampling parameter; and π_{ii} is the conditional probability of E_i given E_i on the previous trial.

According to Equation 1, if the values of N_c and j are small, the asymptotic probability of a repetition response varies approximately from π_{ii} to π as a function of increases in t, the intertrial interval parameter. If the values of N_c and j are not small, asymptotic probability still varies with t, but the range of variation is curtailed in the direction of π.

Equation 1 refers only to the asymptotic probability of repetition responses following one previous reinforcing event. If a nonnegligible value of $(1-j)^t$, the quantity representing the stimulus trace, spans more than one intertrial interval, Equation 1 must be expanded to handle traces from more remote events. Such an expansion has been done (Witte, 1959), and henceforth it will be referred to as Equation 1a.

Equation 1a predicts the asymptotic probability of repetition responses as a function of the number of available traces, x, where x ranges in value from 0 to n. When predicted values are plotted against x, the form of the resulting function depends primarily on the relative value of reinforcement and the particular sequence of events.[2] Thus, for example, if $\pi_{ii}>\pi$, an increasing monotonic function occurs when the n traces originate from a sequence of n

---

[1] This investigation was conducted at Cornell College. It was supported in part by a research grant (M-2848A) from the National Institute of Mental Health, United States Public Health Service, and in part by a grant (G-7962) from the National Science Foundation. Paul Dude, Brian Klepinger, James Overly, Jenny Waychoff, and Robert Beckmen collected most of the data. The final draft of this paper was written during the tenure of a postdoctoral fellowship (MF-11219) from the National Institute of Mental Health, United States Public Health Service.

[2] In some cases, the form of the function also depends on the values of θ and N_c/N, but this complication is ignored for simplicity.

similar events ($E_iE_iE_iE_i\cdots$), but a decreasing cyclical function occurs when the n traces originate from a sequence of n alternating events ($E_jE_iE_jE_i\cdots$). Although too numerous for itemization, all possible combinations of reinforcement values and event sequences yield functions which are determined uniquely by Equation 1a.

Regardless of the value of x, Equation 1a generates essentially the same predictions as Equation 1 for the asymptotic probability of repetition responses as a function of t. To obtain a single, composite prediction from Equation 1a for a particular value of t, one can find the expected predicted value over all possible sequences of n events. As this procedure is repeated for larger values of t, the expected value approaches π in agreement with the trend predicted by Equation 1 for the special case where $x = 1$.

Several studies lend qualitative support to the predictions of the trace-fluctuation model for the massed-trial case. Both Engler (1958) and Anderson (1960) used human *S*s in a verbal conditioning design involving sequentially dependent rewards. In general, these *E*s found probability matching between conditional (and unconditional) response and reward probabilities. After analyzing heretofore unpublished data, Engler (1958) also reported that *S*s in the study by Hake and Hyman (1953) matched conditional response and reward probabilities.

Anderson (1960) tested two models from statistical learning theory, including one proposed by Burke and Estes (1957) which has many properties in common with the present theory. He found large and consistent terminal differences between predicted and obtained values when $\pi_{ii} < .40$. Anderson also suggested that neither model which he tested was able to account for the observed proportions of repetition responses following various sequences of reinforcing events.

The trace-fluctuation model was tested explicitly in a T maze study with rats (Witte, 1959). As trial spacing was increased from 8 sec. to 1 hr., the terminal proportions of repetition responses approached π in accordance with theoretical expectations. However, the quantitative fit between theory and data was not conspicuously successful for either the above analysis or a more detailed dealing with behavior following various reward sequences.

These discrepancies may have been caused by the selective elimination of *S*s with position preferences. The remainder of this paper describes an attempt to test the theory under conditions designed to minimize position preferences or, at least, to avoid the selective elimination of *S*s.

METHOD

Design.—The major independent variables were two symmetrical, sequentially dependent reward schedules and four intertrial intervals. Except for the 24-hr. interval between experimental days, all *S*s worked under a 12-sec. interval during the 360-trial "acquisition" series. Then, presumably after a behavioral asymptote had been reached, *S*s under a given reward schedule were divided at random into four groups. These groups underwent 50 "asymptotic" trials in each of four intertrial intervals in accordance with a systematically selected latin square design. The intervals were 12 sec., 32 sec., 3 min., and 30 min.

A correction procedure insured that each *S* was rewarded equally often on both sides of the T maze. Approximately three-fifths of the *S*s worked under an "alternation" reward schedule in which E_i followed E_i 10% of the time, while E_i followed E_j 90% of the time. The remaining *S*s worked under a "repetition" reward schedule in which E_i followed E_i 90% of the time, while E_i followed E_j 10% of the time. A reward schedule was predetermined separately for each *S* over an initial block of 60 trials and subsequent blocks of 50 trials. These schedules were constructed from tables of random numbers with the restriction that the sequence reflected, within a margin of one event, the unconditional and conditional frequencies specified in the experimental design.

Several features of the design merit additional comment. Previous findings (Witte, 1959) indicated that position preferences occurred most frequently among *S*s under highly distributed trials and among *S*s under alternation reward schedules. In an attempt to minimize the former effect, trials were massed during the acquisition series for all *S*s, and to compensate for the latter effect, a majority of *S*s were assigned to the alternation schedule. Notwithstanding these precautions, a large incidence of position preferences occurred among the alternation *S*s during acquisition. Accordingly, only the repetition *S*s were rotated through the four intertrial intervals of the asymptotic series.

Apparatus.—The apparatus, a simple T maze, has been described elsewhere (Witte, 1959). In the present study, a heavy cardboard partition, extending from the maze table to the ceiling of the experimental room, separated the two branches of the maze from a desk chair which straddled the stem of the maze. The partition had two 8½ × 18½ in. horizontal apertures which permitted *E* to observe *S*s' movements in the branches. Covered with light blue gelatin paper (#129, Rosco Lab.), the apertures functioned as one-way screens, since the only light sources in the room were three 100-w. bulbs on the branch side of the partition.

Procedure.—A 22½-hr. food deprivation schedule was enforced throughout the experiment. All groups of *S*s participated in a counterbalanced design yielding a mean deprivation of 20 hr. at running time.

After a 12-day training period, described elsewhere (Witte, 1959), trials in the T maze were begun. An additional 12 days elapsed before *S*s' running times permitted the desired five trials per day. On each trial, after *S* had been placed in the startbox, the gates at the startbox and at the choice point were raised. When *S*'s entire body had entered a given branch alleyway, these gates were lowered, and one of the goalbox doors was unlocked according to a predetermined schedule. If *S* ran to the side containing the locked door, he was forced to return to the choice point and proceed to the other side. After *S* had eaten a .045-gm. food tablet (P. J. Noyes Co.), he was removed from the goalbox and placed in an individual detention cage for a prescribed period of time before receiving the next trial.

By definition, the intertrial interval began when *E* removed the lid of the goalbox, and it ended when *S* reached the choice point on the next trial. Since the time required by most *S*s to get from the startbox to the choice point stabilized at about 2 sec., the intertrial interval could be specified within a small margin of error. Because of the correction procedure, the size of the *intratrial* interval was relatively variable, but it averaged about 8 sec.

A pilot study showed that a single *E* tended to be harried, and as a result, probably cued the location of the reward on some trials. In order to minimize inadvertent cues, *E*s worked in pairs. One *E* handled *S*s and operated the gates in the maze stem, while the other *E*, sitting in the desk chair described above, recorded *S*s' responses, controlled the goalbox doors, and timed the intertrial intervals. All *E*s were closely supervised, and special meetings were conducted to insure a uniform treatment.

Subjects.—The *S*s were 64 naive hooded rats between 5 and 6 mo. of age at the beginning of the experiment. Three fatalities occurred during the experiment, and 6 *S*s were discarded because they consistently refused to run in the maze.

RESULTS

The response was defined as the first choice of *S* on each trial. Left and right turns at the choice point were designated as A_1 and A_2 responses; food rewards in the left and right goalboxes were identified with E_1 and E_2 events. Unless noted otherwise, all analyses were based on responses from Trials 2–5 of each experimental day, and terminal estimates of response probability for the acquisition series were derived from the last 20 days of the series.

Unconditional probabilities.—According to statistical learning theory (Estes, 1954), the terminal A_1-response proportions of individual *S*s should approximate a binomial distribution about π. To test the group mean or π prediction, terminal estimates of .51 ($s_p = .046$) and .52 ($s_p = .014$) were obtained for the 31 alternation *S*s and the 24 repetition *S*s during the acquisition series, and an estimate of .48 ($s_p = .018$) was obtained for the repetition *S*s during the 40-day asymptotic series. Obviously, none of these estimates deviated significantly from the predicted value of $\pi = .50$.

To test the binomial prediction, the expected distribution of A_1-response frequencies was determined for *S*s within a given group by referring to tables of the binomial distribution (Office of Chief of Ordnance, 1952), and the discrepancies between obtained and expected frequencies were evaluated by computing *D*, the statistic associated with the Kolmogorov-Smirnov test for goodness of fit (Siegel, 1956).

The terminal distribution of A_1 responses for the alternation *S*s during acquisition departed significantly ($D = .39$, $P < .01$) from binomiality. A further analysis revealed that nonbinomiality was caused primarily by the large incidence of *S*s with position preferences. More specifically, 22 of the alternation *S*s fell outside the 99% acceptance interval for .50. The stability of these preferences was indicated by their early appearance during the 80-day acquisition series. For example, over one-half of the preferences appeared during the first 30 days. It should be noted, however, that there was no consistent preference among the 22 *S*s for a particular side of the maze. This was also the case for all other subgroups of *S*s with position preferences in the present study.

Although the repetition *S*s also departed significantly ($D = .37$ and $.38$, $P < .01$) from binomiality during the terminal phase of acquisition and the first asymptotic rotation, they approximated ($.15 \leq D \leq .17$, $P > .20$) the predicted distribution during each of the last three asymptotic rotations. The two departures from binomiality by the repetition *S*s were due mainly to bimodality and skewness, rather than to a large incidence of position preferences. Thus, in the acquisition analysis, 3 *S*s barely exceeded the 99% acceptance interval for .50, their preferences appearing only during the final 20–30 days. In the asymptotic analyses, a total of 6 *S*s fell outside the acceptance interval during at least one rotation. Typically, these preferences lasted only for one rotation, and they were distributed evenly among the four rotations and the four intertrial intervals.

The π and binomial predictions were also applied to A_1 responses on first daily trials during the last 20 days of acquisition for both groups, and during the 40 asymptotic days for the repetition *S*s. In all three cases, although the resulting mean estimates deviated insignificantly from π, the distributions departed significantly ($.32 \leq D \leq .42$, $P < .01$) from binomiality. In every case, the departure from binomiality was caused primarily by the large number of position preferences: 17 for the alternation *S*s, and 15 and 12 for the repetition *S*s during the acquisition and asymptotic series. These results for the repetition *S*s, when compared with the previous paragraph, indicate less stereotypy during later daily trials.

Conditional probabilities.—In this section, measures of conditional response probability reflect the performance of all *S*s within a given group. Unless noted otherwise, it may be inferred that no appreciable difference existed between the overall measure and that for any subgroup of *S*s without position preferences.

In addition to minimizing position preferences, the acquisition series was introduced to produce a stable proportion of repetition responses before *E* manipulated the size of the inter-

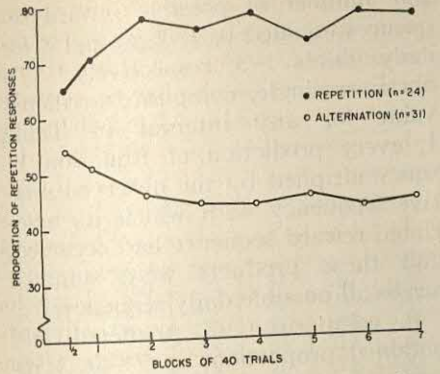


FIG. 1. Obtained proportion of repetition responses during acquisition.

trial interval. The acquisition curves for repetition responses are shown in Fig. 1. Both groups fell above the 50% level in the initial trials. Following rapid changes over Trials 21–80, i.e., Days 13–28, the alternation *S*s approached a terminal value of .44 (s_p = .016), and the repetition *S*s stabilized at .78 (s_p = .013). The curve for the subgroup of 9 alternation *S*s without position preferences was lower by .03 than the alternation curve of Fig. 1. Because of the large incidence of position preferences among alternation *S*s, no further results will be given for these *S*s.

During the asymptotic series, as will be recalled, the repetition *S*s spent 10 days under each of the four intertrial intervals. According to the theory, the asymptotic proportion of repetition responses for these *S*s should approximate π as trial spacing increased. The obtained proportions, shown in Table 1, provide qualitative support for this prediction.

An analysis of variance was performed over the obtained proportions summarized in the first four intervals of Table 1. The effects associated with rotation period, order of presentation, and within-*S* interaction were not significant, while, not unexpectedly, the intertrial interval effect was significant ($F = 14.06$, $df = 3/6$, $P < .001$). A further analysis of the latter effect indicated that all paired comparisons, except the one between 12- and 32-sec. intervals, reached at least the .05 level of significance. A final analysis compared performance under the 12-sec. interval during the asymptotic series and during the terminal portion of the acquisition series. The resulting mean difference of .03 was not significant, as expected, assuming that a behavioral asymptote had been reached during the acquisition series.

Values of six parameters were required before the predicted proportions in Table 1 could be derived from Equation 1a. Three of these parameters, including the mean intratrial interval parameter, c, which appears in Equation 1a, were evaluated prior to an inspection of the repetition response data. Their values were $\pi_{ii} = .90$, $\pi = .50$, and $c = 8$.

The remaining parameters were estimated from the repetition response data after a simplifying assumption had been made. According to the theory, if no traces from previous reinforcing events are present on a given trial, only the set of common stimulus elements remains to be

TABLE 1

PROBABILITY OF REPETITION RESPONSES AS A FUNCTION OF INTERTRIAL INTERVAL: ASYMPTOTIC SERIES

Probability	Intertrial Interval				
	12 Sec.	32 Sec.	3 Min.	30 Min.	24 Hr.
Predicted	.75	.74	.73	.60	.52
Obtained	.75	.73	.66	.61	.53[a]
SE[b]	.019	.021	.022	.017	.016

[a] Based on first daily responses during the asymptotic series.

[b] Standard error of obtained probability.

sampled, and therefore, the asymptotic proportion of repetition responses on that trial should equal π (plus a small "correction" term). The proximity of π and .53, the obtained proportion of repetition responses on all first daily trials during the asymptotic series, implies that no traces spanned the 24-hr. interval between experimental days. This inference receives additional support from an estimate of .52 for first daily trials during the last 20 days of the acquisition series.

The simplifying assumption had several immediate consequences. First, the difference between observed proportions of repetition and alternation responses on first daily trials provided a sequential estimate of θ (Burke & Estes, 1957; Estes & Straughan, 1954). A sequential estimate of .04 for θ was obtained from the above proportion for the acquisition series, rather than its asymptotic counterpart, in order to avoid estimating parameters with to-be-predicted data.

Second, it could be inferred that not more than one trace was present on Trial 2 of each experimental day. Accordingly, Equation 1, instead of the more elaborate Equation 1a, was used to obtain estimates of .001 for j, the rate of decay, and .95 for N_c/N, the ratio of common elements to trace elements. These estimates represented solutions for a pair of expressions involving Equation 1 and two data points: the proportion of repetition responses on Trial 2 during the last 20 days of acquisition and the proportion on Trial 2 during the 10 asymptotic days which *S*s spent under the 30-min. intertrial interval.

Given the value of the intertrial interval, as well as the values of the six parameters, Equation 1a predicted the proportion of repetition responses on a particular daily trial as a function of each possible sequence of rewards. The number of possible reward sequences equaled 0, 1, 2, 4, and 8 for daily Trials 1–5, respectively.[3] To obtain a single, composite predicted value for any interval in Table 1, every prediction of Equation 1a was multiplied by the observed relative frequency with which its associated reward sequence had occurred, and these products were summed across all possible daily sequences.

A comparison of predicted and obtained proportions in Table 1 reveals a fairly large discrepancy at the 3-min. interval ($t = 3.0$, $P < .01$). Otherwise, the fit was excellent, particularly since only data from the 30-min. interval were used in estimating theoretical parameters.

Sequence probabilities.—An analysis of repetition responses as a function of daily reward sequences permits a more incisive, "fine-grained" test of Equation 1a. To facilitate this analysis, the immediately preceding reward, relative to which the repetition response was defined, is denoted by 1, regardless of whether it was an E_1 or E_2 event. More remote events are designated 1 or 2, depending on whether or not they differed from the referent event. Thus, for example, the symbol R(221) represents the proportion of repetition responses on Trial 4, given the referent event on Trial 3 and the alternative event on Trials 2 and 1. Note that the numbers in the above parentheses read from right to left. The present tabulational procedure differs from those used by Anderson (1960) and Nicks (1959) in that the data from each trial are being considered separately.

Table 2 shows the 10 most frequently occurring R values for the repetition *S*s, ranked according to predictions from Equation 1a. Except for an occasional reversal, nota-

---

[3] For example, the two possible reward sequences on Trial 3 were E_iE_i and E_jE_i.

TABLE 2

Probability of Repetition Responses as a Function of Reinforcement Sequences: Repetition *S*s

Sequence[a]	*N*[b]	Blocks of 40 Trials										
		Acquisition							Asymptotic Interval			
		1	2	3	4	5	6	7	12 Sec.	32 Sec.	180 Sec.	1800 Sec.
R(1111)	160	.80	.84	.86	.90	.85	.89	.91	.88	.85	.73	.71
R(111)	184	.83	.85	.89	.85	.82	.88	.88	.88	.83	.70	.67
R(11)	213	.75	.87	.78	.87	.77	.86	.84	.85	.79	.77	.65
R(1)[c]	242	.65	.68	.73	.69	.66	.73	.70	.68	.67	.67	.57
R(2111)	19	.78	.76	.74	.82	.76	.74	.68	.67	.67	.62	.62
R(211)	21	.62	.62	.56	.59	.65	.50	.63	.39	.61	.41	.50
R(2211)	23	.68	.71	.50	.65	.53	.52	.47	.50	.25	.31	.62
R(21)	22	.46	.47	.40	.32	.50	.43	.37	.26	.33	.45	.36
R(221)	22	.55	.41	.22	.26	.35	.40	.32	.23	.31	.30	.35
R(2221)	22	.33	.26	.24	.35	.34	.20	.22	.07	.45	.26	.23
Summary *d*[d]		12	10	6	10	2	4	0	2	6	10	4

[a] Ranked from highest to lowest in terms of predicted values.

[b] Mean total frequency of occurrence per block of 40 trials.

[c] The theory predicts R(2111) > R(1) for the 180-sec. interval and R(2111) > R(211) > R(1) for the 1800-sec. interval.

[d] The letter *d* refers to the sum of the absolute discrepancies between predicted and obtained ranks, and it has a maximum value of 50.

bly at the fourth block of acquisition trials and at the 3-min. interval, the qualitative fit between predicted and obtained ranks improved progressively and was eventually quite good.

The large and consistent obtained differences between selected pairs of R values, e.g., R(1111) and R(2111), implies that *S*s' behavior was influenced by the more remote daily

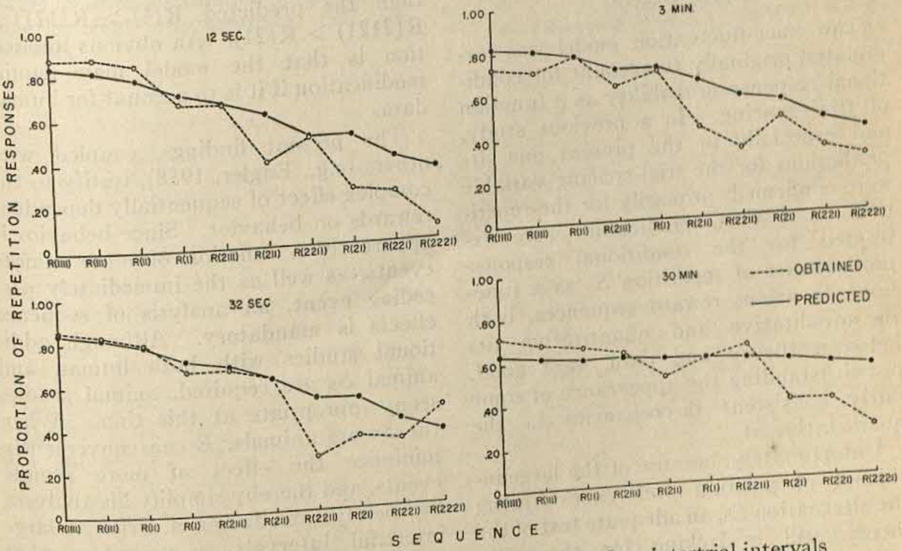


FIG. 2. Predicted and obtained R values during the four intertrial intervals of the asymptotic series.

rewards. With minor exceptions, however, the rank distributions for the four most frequently occurring R values in Table 2 failed to reflect Jarvik's (1951) negative recency effect. Instead, in agreement with the theory, repetition responses occurred more frequently as sequences of similar events increased in length.

The asymptotic R values for the repetition *S*s were also compared with the set of exact predictions which had been computed previously to determine the expected proportions in Table 1. Inspection of Fig. 2 reveals that Equation 1a consistently overestimated R values for dissimilar events ($P < .01$), and underestimated R values for similar events, although the latter tendency was not significant, according to a binomial test. Generally speaking, the discrepancies were smaller for similar events than for the less reliable dissimilar events. The best fit between theory and data occurred at the 32-sec. interval, while the poorest fit occurred at the 3-min. interval, where the theory overestimated all R values except one.

Discussion

The trace-fluctuation model was formulated originally to account for conditional response probability as a function of trial spacing. In a previous study, and especially in the present one, its predictions for the trial-spacing variable were confirmed, primarily for the repetition *S*s. When predictions were extracted for the conditional response probabilities of repetition *S*s as a function of various reward sequences, both the qualitative and quantitative fits between theory and data were good, notwithstanding the appearance of some fairly consistent discrepancies in the quantitative fit.

Unfortunately, because of the large incidence of position preferences among the alternation *S*s, an adequate test of the theory still is lacking for these *S*s. During massed trials, substantially more position preferences appeared among the alternation *S*s in the present study than among those in a previous study (Witte, 1959). Although the two studies are not strictly comparable because of differences in procedure and *S*s, an important factor may have been the 4-sec. increase in trial spacing for the present study. This conjecture is strengthened by the fact that in the earlier study, position preferences appeared among almost all of the alternation *S*s under a 30-sec. intertrial interval. Thus, to minimize position preferences among alternation *S*s in future studies, it seems advisable to train *S*s under trials which are as highly massed as possible.

The theory did not fare so well when it was checked against the sequence data reported by Anderson (1960) for nine groups of human *S*s, each of which received a different sequentially dependent reward schedule. In general, when his R values for Trials 101–300 were ranked according to the qualitative predictions of the theory, the agreement between predicted and obtained ranks was rather poor, particularly for sequences of dissimilar events. For example, when $\pi_{ii} = .70$, Anderson found $R(1) > R(21) > R(121) > R(2121)$ rather than the predicted $R(1) > R(121) > R(2121) > R(21)$. An obvious implication is that the model may require modification if it is to account for human data.

The present findings, coupled with others (e.g., Engler, 1958), testify to the complex effect of sequentially dependent rewards on behavior. Since behavior is apparently a function of more remote events, as well as the immediately preceding event, an analysis of sequence effects is mandatory. Although additional studies with both human and animal *S*s are required, animal studies seem appropriate at this time. When the *S*s are animals, *E* can conveniently minimize the effect of more remote events, and thereby simplify his analysis, by merely introducing a series of large intertrial intervals, a procedure that tends to be impractical for human *S*s.

This advantage should outweigh the inconvenience of position preferences.

SUMMARY

A model derived from statistical learning theory was tested with rats in a T maze. The major independent variables were two sequentially dependent reward schedules and four intertrial intervals. After preliminary training under massed trials, the *S*s under a given reward schedule were rotated through each intertrial interval in accordance with a latin square design.

During preliminary training, position preferences were minimal for *S*s under the "repetition" schedule, but not for *S*s under the "alternation" schedule. Subsequently, only the repetition *S*s were rotated through the intertrial intervals. The conditional response probabilities for these *S*s declined as a function of trial spacing. In general, the theory successfully predicted changes in conditional response probability as a function of both trial spacing and various sequences of rewards.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

ENGLER, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

ESTES, W. K. Individual behavior in uncertain situations: An interpretation in terms of statistical association theory. In R. A. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, 1954. Pp. 141–149.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

OFFICE OF CHIEF OF ORDNANCE. Tables of the cumulative binomial probabilities. *Ordn. Corps Pamphl.*, 1952, No. ORDP 20–1.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

WITTE, R. S. A stimulus-trace hypothesis for statistical learning theory. *J. exp. Psychol.*, 1959, **57**, 273–283.

(Received August 4, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 448–455

AN APPLICATION OF STIMULUS SAMPLING THEORY TO SUMMATED GENERALIZATION[1]

TERESA S. CARTERETTE[2]

Indiana University

The general paradigm for experiments demonstrating summated generalization used, for example, by Kalish and Guttman (1957) with pigeons and by Bilodeau, Brown, and Meryman (1956), with human *S*s, is to reinforce a response at two points in a stimulus dimension and, then, to test for generalization along the dimension. The present study represents an attempt to go beyond Bilodeau, Brown, and Meryman by obtaining a more precise determination of the form of the summated gradient for human *S*s and by relating its specific properties to the separation between training stimuli. In addition, the experiment tests a mathematical model for stimulus generalization along a substitutive dimension (Stevens, 1957) derived from the Estes-Burke probabilistic theory of behavior.[3]

Basic to the model is the assumption that, for equal intervals along the dimension, $P(Y) = f(\Delta S)$ is a linear function with negative slope. The dependent variable, $P(Y)$, is the probability of the conditioned response; and (ΔS) is a change along the physical scale of the training stimulus. While Guttman and Kalish's (1956) data for pigeons responding to the wavelength of light generally support the assumption of linearity, human *S*s responding to the position of a light (Brown, Bilodeau, & Baron, 1951) showed such a great drop in response level at the test stimulus immediately adjacent to the training stimulus that clearly most of the gradient was not being adequately sampled. To avoid this high degree of discriminability, the position of the light in the present situation was varied over a continuous field rather than with discrete lamps as in the earlier experiment.

The rationale underlying the simple generalization model and the predictions for summated generalization are elaborated formally in a later section. Briefly, each stimulus in the simple gradient is associated with a set of elements; and some of the elements in each of the sets corresponding to each of the generalized stimuli are assumed to be also present in the set corresponding to the training stimulus. Since the theory relates response probability directly to the proportion of stimulus elements conditioned to a response, the extent to which a test stimulus elicits the reference response reflects the proportion of elements which it shares with the training stimulus. In the summated generalization case where the same response is conditioned to two stimuli in the dimension, response probability should be determined by the proportion of elements which a given test stimulus shares with the two training stimuli.

---

[1] This paper is based on a PhD dissertation submitted to the Graduate School of Indiana University. The work was executed while the author held a National Institutes of Health Predoctoral Research Fellowship under the sponsorship of W. K. Estes who provided valuable counsel at all stages of the project.

[2] Now at Institute for Psychological Research, Tufts University.

[3] A preliminary presentation of the application of stimulus sampling theory to stimulus generalization phenomena has appeared in mimeographed lectures given by W. K. Estes before the Summer Institute in Social Sciences for College Teachers of Mathematics held at Stanford University in the summer of 1957.

Application of the assumptions of the set theoretical model (Burke & Estes, 1957) to the simple generalization data obtained in Part I of the experiment provided parameter estimates for the a priori predictions tested in Part II, a summated generalization experiment. Predictions pertained to three aspects of the summated generalization gradients. It was expected that: (*a*) the height at the two training stimuli would be identical to that obtained in the simple generalization experiment, (*b*) midway between training stimuli, the height of the gradients would decrease as the separation between training stimuli increased, and (*c*) the slopes of the outer segments of the summated gradients would be identical for all separation groups.

METHOD

Subjects

The *S*s were 240 Indiana University undergraduates. Forty-eight provided the data for the simple generalization gradients; and the remainder provided the summated generalization data.

Apparatus

The apparatus enabled *E* to present a $2\frac{3}{4}$-in. square of light for 1 sec. at different positions along a horizontal line. The *S* faced a black panel which supported a 5 in. high by 44 in. wide field of sand blasted plate glass on which the stimulus light appeared. The walls and ceiling of the room were black; and ambient light was provided by a 7.5-w. frosted ceiling bulb situated directly above *S*.

Design

Simple generalization.—For Group C (Center) the training stimulus was located exactly midway between the left and right limits of the field; and for Group Q (Quarter), it was one-fourth the distance from its left border. The remaining stimuli were to the right of the training stimulus and covered a 3.5-in. range in $\frac{1}{2}$-in. steps.

Summated generalization.—For the Center groups the left-hand training stimulus was at the midway position and for the Quarter groups, at the point one-fourth the distance from the left limit of the stimulus field. The second training stimulus was four $\frac{1}{2}$-in. steps to the right of the first for the T_4 groups; and six and eight steps for the T_6 and T_8 groups, respectively. These separation values were intended to realize two types of situations. In the T_4 groups the stimulus sets associated with the two training stimuli should have a proportion of elements in common. For the T_6 and T_8 groups, these sets should be discrete; but the proportion of elements which the central stimuli share with the training stimuli should differ.

Procedure

Simple generalization.—Each *S* was tested separately in a 40-min. session with 10 min. for dark adaptation and instructions which covered the following points: (*a*) There would be two types of trials: trials on which *E* would announce that the stimulus was to appear in the standard position, and test trials on which the unannounced stimulus might be in the standard or in some different position. (*b*) The stimulus duration would be 1 sec.; and on test trials *S* must decide whether the light had been in the standard or in some different position. (*c*) The two keys in front of *S* were to be used in communicating judgments to *E*. The training stimulus, presented once at the outset of every block of trials, was followed by 8 test trials in which the standard and seven other stimuli each appeared, these blocks continuing to a total of 180 trials. The order of test stimuli within each block was randomized separately for each *S*. To indicate the positive response, half the *S*s were instructed to use the right-hand key and half, the left key. Approximately 9 sec. elapsed between stimulus presentations.

Summated generalization.—Following dark adaptation and instruction, *S* was exposed to 210 trials. Instructions were altered as follows: (*a*) There would be three types of trials: trials on which *E* would announce the stimulus position as *A*, trials on which *E* would announce it as *B*, and test trials on which the unannounced stimulus might be in *A*, or in *B*, or in some different position. (*b*) On test trials if *S* believed that the light was in *A* or *B*, he was to press the right-hand key; and if he believed that it was in neither position, he was to press the left-hand key. Within each block of 21 trials there were 4 training trials. The two training stimuli each appeared twice, once at the beginning and once midway in the block. The order of test stimuli was randomized with the restriction that each stimulus, including *A* and *B*, appear only once in the 17 test trials.

RESULTS

Simple generalization.—Gradients for Groups C and Q were linear over a 2.5-in. range. Comparison of performance for the first 10 blocks of trials with that for the second 10 blocks did not reveal the inhibitory effects observed by Bilodeau, Brown, and Meryman (1956) who found that as test trials progressed, *S*s produced fewer responses to generalized stimuli. A straight line fitted to the combined data by the method of least squares had a slope of −.16 and its intercept at .82.

Guttman and Kalish (1956) found that gradients for pigeons pecking a key in the presence of different wavelengths of light became flatter as the total number of generalized responses decreased. In the present case three gradients for *S*s grouped on the basis of the total number of positive responses did not have different slopes.

Summated generalization.—Figure 1 presents the mean proportion of positive responses at each test position over the 10 blocks of trials. Response level along the center segments decreases as the distance between training stimuli increases. In the T_4 groups these data points are slightly above those at the training stimuli. For the T_6 groups there is virtually no difference between the training stimuli and the intermediate positions; and in the T_8 groups a decline in response level is clear.

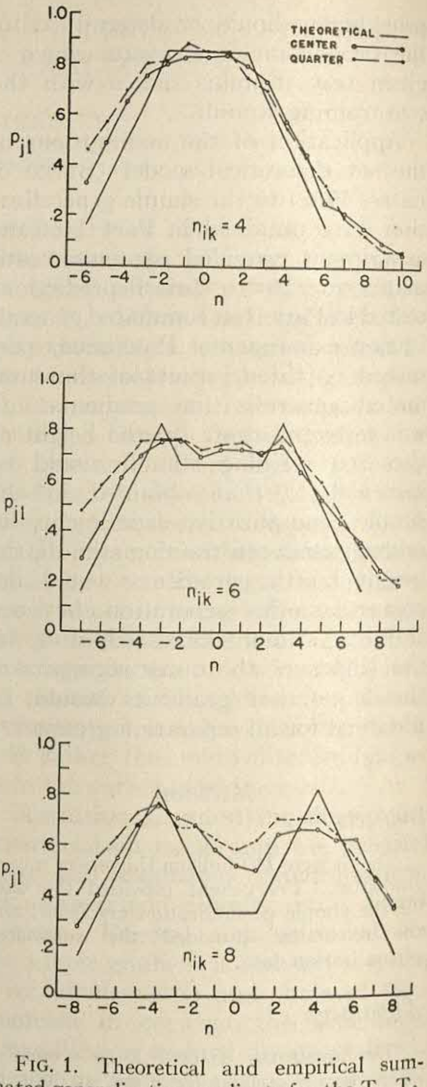


FIG. 1. Theoretical and empirical summated generalization gradients for the T_4, T_6, and T_8 groups. (Zero corresponds to the stimulus exactly midway between training stimuli; and unit changes represent ½-in. steps along the physical scale.)

In the absence of normal distributions, nonparametric tests were applied to the data for the entire session. To test the hypothesis that response level at the midway stimulus decreases with increasing separation between training stimuli, each *S*'s mean score at the midway position was subtracted from the mean of his pooled scores at the training stimuli. The Kruskal-Wallis one-way analysis of variance statistic, H, is significant at the .05 level with 2 *df* for both Group C and Group Q comparisons. The t's comparing the mean proportion of positive responses to the training stimuli with that at the midway stimulus are significant at the .05 level with 62 *df* for the QT_4 and for the T_8 groups only.

The outer segments of all gradients are linear over four stimulus positions. Two measures were used in comparing the slopes of corresponding outer segments: (*a*) each *S*'s mean proportion of positive responses over the four outer stimuli adjacent to a given training stimulus, and (*b*) the difference between each *S*'s pooled mean proportion of positive responses at the first two and second two of these stimuli. The H for Measure 1 on the left side and that for Measure 2 on the right side for Group C gradients are significant at the .05 level with 2 *df*. Although the corresponding outer segments of the gradients for different separation groups are not far from parallel, the possibility of some significant local deviations cannot be ruled out by the present analysis.

Because Group Q gradients are asymmetrical, an effect probably due to the presence of the left border of the frame near this part of the field, tests for symmetry were restricted to Group C data. The Mann-Whitney U is significant at the .05 level for Measure 2 in Group CT_6. The data for Group CT_8 indicate that departures from symmetry in the Center groups may be due to the use of the center position for one of the training stimuli; but any general statement concerning symmetry irrespective of separation between training stimuli is necessarily tenuous.

Although gradients for the first and second halves of the experimental session show no systematic change in response strength for the Center groups, the response level between training stimuli tends to be lower over the second half of the Quarter groups.

THEORETICAL ANALYSES

Formal Theory

Correspondence between psychological variables and mathematical sets is defined by the application of the Burke-Estes (1957) component model for stimulus variables. A mathematical set, S_i, is associated with each stimulus. Sets $S_0, S_1, \cdots S_k$ denote $k+1$ distinguishable stimuli; and the overlap between S_i and S_j represents the similarity between stimuli i and j. The number of elements in any set S_i is N; and the number which S_i and S_j have in common is N_{ij}. Behaviors available to S are represented by A_1 and A_2, the mutually exclusive and exhaustive responses formally distinguished in the experiment; and it is assumed that each element in a set is connected to one of these responses so that, given stimulus i, the probability of A_1 is the proportion of elements in i connected to A_1 and sampled by S on that trial.

The relationship between stimulus and response variables, conceptualized in this manner, has been elaborated in theoretical papers by Burke and Estes (Burke & Estes, 1957; Estes & Burke, 1953). The probability of an A_1 when i is presented is expressed by:

$$p_i = \frac{N(C_i)}{N} \qquad [1]$$

where $N(C_i)$ is the number of elements in S_i connected to A_1.

Generalization of response A_1 to stimulus j, given i as the training stimulus, may be expressed as the sum of the proportion of elements common to S_i and S_j which are conditioned to A_1 by virtue of training on i and the proportion of elements in S_j only which are connected to A_1 independently of training on i:

$$p_j = \frac{N_{ij}p_i + (N - N_{ij})\rho_j}{N} \qquad [2]$$

where ρ_j, a constant for all j, is the probability of an A_1 given j as the stimulus prior to training on i. Substituting w_{ij} for (N_{ij}/N) and rearranging terms, a more convenient form of Equation 2 is obtained:

$$p_j = w_{ij}(p_i - \rho_j) + \rho_j \qquad [3]$$

The range over which generalization will occur is determined by the extent of stimulus overlap between the training stimulus and other stimuli in the series. Given training with reinforcement of response A_1 to stimulus i and nonreinforcement elsewhere along the dimension, or in the case of a voluntary response, following instructions to make response A_1 to stimulus i only, it may be assumed that ρ_j equals zero for $j \neq i$. Equation 3 becomes:

$$p_{(i+n)} = w_{i(i+n)}p_i$$

where n is the number of equal intervals on the physical dimension separating i and j; and it may be restated in the form:

$$\begin{aligned} p_{(i+n)} &= p_i - p_i(1 - w_{i(i+n)}) \\ &= p_i - n(1 - w_{i(i+1)})p_i \qquad [4] \end{aligned}$$

Equation 4 follows from the fact that stimuli equidistant on the dimension have equal overlaps, so that $w_{i(i+1)} = w_{(i+1)(i+2)} \cdots$ etc. Hence, $n(1 - w_{i(i+1)})$ is the proportion of new elements introduced for a stimulus n steps away from i. If n exceeds $1/(1 - w_{i(i+1)})$, Equation 4 no longer applies.

This equation may be used for predicting generalization along a substitutive dimension. A family of linear gradients may be plotted with p_i as a parameter depending only on learning and n as the independent variable. For experimental convenience it is desirable to determine equal intervals on the physical dimension corresponding to equal decrements in p_i by obtaining a simple generalization gradient empirically. Over the range of physical values for which the gradient fits a straight line, the overlap between i and successive equally separated stimuli may be assumed to decrease by a constant proportion; and the parameters p_i and $w_{i(i+1)}$ may then be estimated from the data and used to test the predictive power of the model in an entirely new situation.

For summated generalization with training on i and k, p_j is the proportion of conditioned elements which j shares with i and k:

$$\begin{aligned} p_j &= \frac{N_{ij}p_i + N_{jk}p_k - N_{ik}p_ip_k}{N} \\ &= w_{ij}p_i + w_{jk}p_k - w_{ik}p_ip_k \\ &= p_i[1 - n_{ij}(1 - a)] \\ &\quad + p_k[1 - n_{jk}(1 - a)] \\ &\quad - p_ip_k[1 - n_{ik}(1 - a)] \qquad [5] \end{aligned}$$

For simplicity in notation $w_{i(i+1)}$ has been replaced with a. The quantity $[1 - n_{ik}(1 - a)]$ is the proportion of overlap between the two training stimuli; n_{ik}, the number of steps between i and k; n_{ij} and n_{jk}, the numbers of steps between a test stimulus j, and i and k, respectively. The third term on the right-hand side corrects for the fact that when training stimuli overlap, the sets associated with stimuli between i and k contain elements common to S_i and S_k which have been included twice in the sum of the first two terms.

Two types of relations are determined by n_{ik}. For n_{ik} between zero and $1/(1 - a)$ the training stimuli overlap; and their intersection must be taken into account. If $n_{ik} \geqslant 1/(1 - a)$, there is no such overlap; and the third term on the right side of Equation 5 is zero. Beyond certain points in the dimension, j will cease to overlap with part of the intersection of i and k, or with either or both training stimuli; and restrictions must be placed on n_{ij} and n_{jk} in both instances. In the present experiment, values of n_{ik} were selected to represent these two types of relations. Given the overlap parameter estimated from the data, training stimuli which are four steps apart, the value of n_{ik} for the T_4 groups, should overlap. Sets associated with stimuli separated by more than four steps should be discrete; and the T_6 and T_8 groups are intended to represent special cases of this type.

Prediction 1.—For all $n_{ik} \leqslant 1/(1 - a)$ the sets associated with test stimuli $j = (i + 1) \cdots (i + n - 1)$, which lie between i and k, are contained in the union of S_i and S_k; and $p_i = p_j = \cdots p_k$. The solid line in the top panel of Fig. 1 presents this type of gradient, the condition expected to obtain for Groups T_4. It should be noted that when $p_i < 1$, the predicted response strength at the training stimuli following summated generalization training is greater when $n_{ik} < 1/(1 - a)$ than when $n_{ik} = 1/(1 - a)$. Theoretical gradients for Groups T_6 and T_8 are shown in the other two panels of Fig. 1. Since the training stimuli do not overlap, the central test stimuli should contain elements not in i or k; and generalization at this portion should decrease as n_{ik} increases. The flat region should extend over $(n_{ik} + 1)$ points if $n_{ik} < 1/(1 - a)$ and over $\left[\frac{2}{1-a} - n_{ik} + 1\right]$ points if $\frac{1}{1 - a} \leqslant n_{ik} < \frac{2}{1 - a}$.

Prediction 2.—Corresponding outer segments of the gradients should be linear with identical slopes for all n_{ik}. For $n_{ik} < 1/(1 - a)$ and $p_i < 1$, the conditioned overlap between i and k produces a departure from linearity which increases as n_{ik} decreases. Gradients should extend over a range of $\left[n_{ik} - 1 + \frac{2}{1-a}\right]$ points.

Adequacy of Predictions for Summated Generalization

The straight line fitted to the simple generalization gradient provided values of p_i and p_j which were substituted in Equation 4 to estimate a. Theoretical and obtained gradients may be compared in Fig. 1.

Prediction 1.—When $n_{ik} = 4$, the discrepancy between theoretical and empirical points along the central segment is in no instance greater than .03. However, the large and consistent disparity between theoretical and obtained values in Groups T_6 and T_8 at these points indicates that predictions from the model are too low when $n_{ik} > 1/(1 - a)$.

A salient, and complicating, feature is the downward shift in response strength at the training stimuli. Since, according to the model, response strength at the training stimuli may vary without affecting a, new theoretical gradients might be derived with p_i and p_k estimated from the data for one of the summated generalization groups; but this approach would yield even lower values of p_j than those obtained with $p_i = p_k = .818$.

Statements concerning the symmetry of the center segment and the extent of the plateau must be highly tentative since experimental artifacts discussed in a previous section may have introduced considerable distortion into the gradients. The data do suggest that p_j may be constant over the predicted number of intervals in the central region; but there is no clear indication that the remaining points are symmetrical with those for the outer segments.

Prediction 2.—The outcome of statistical tests comparing the slopes of corresponding outer segments and of tests of symmetry make it impossible to generalize unequivocally concerning the relevance of the results to predictions from the model. On the whole, the data support the assumption that for a substitutive dimension, p_j varies linearly over n. Empirical points over the left outer segments, especially for the Quarter group gradients, generally lie above the predicted values, a disparity which may possibly be due to the conditioning of extraneous cues to the positive response. The ranges over which the empirical gradients extend exceed predictions.

Estimates of a *from summated generalization data.*—The possibility remains that the stimulus component model may be used to extrapolate across summated generalization situations. An additional estimate of a, obtained by fitting a straight line to the data points for the right hand training stimulus and the six stimuli of the corresponding outer segment of the T_6 groups, yields an overlap parameter equal to .85. In Fig. 2 it can be seen that this estimate results in a better coincidence between theoretical and data points, particularly in the case of Groups T_4 and T_8.

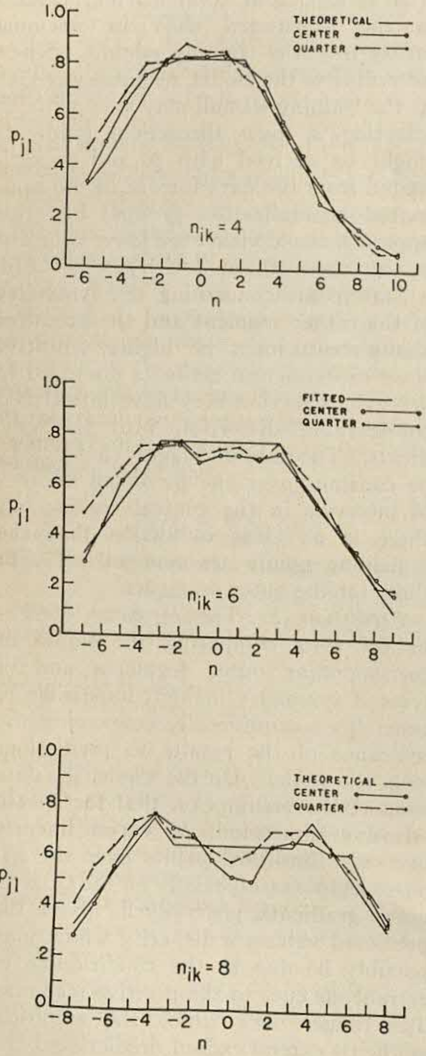


FIG. 2. Theoretical and empirical summated generalization gradients with a estimated from the right hand outer segments of the T_6 groups. (Zero corresponds to the stimulus exactly midway between training stimuli; and unit changes represent ½-in. steps along the physical scale.)

Adequacy of the Theory

In general the two main predictions which follow from the model have been supported. First, the decrease in response level as the separation between training stimuli increases is found in all instances. The second prediction, that corresponding outer slopes of the summated gradients should be parallel for all groups, is well substantiated by the data.

Principal discrepancies between theory and data pertain to three aspects: (*a*) the center dip predicted for the T_6 and T_8 gradients is absent in the T_6 groups and is shallower than expected in the T_8 groups; (*b*) the range of generalization is greater than predicted; and (*c*) the left wings of all gradients, especially in the Quarter groups, are consistently too high. Contrary to prediction and perhaps related to (*a*), the response level for Group QT_4 is significantly higher for the midway stimulus than for tests at the training stimuli. However, it is exactly for the Quarter groups that pronounced asymmetry is present as an experimental artifact. There is the additional possibility that the unequal number of test stimuli on the two outer segments of the T_4 gradients may have interacted with effects of the left border to further increase conditioning to stimuli on the left half of the gradient; consequently, these results remain questionable as legitimate tests of the theory.

If one is mindful of the fact that the parameters used in generating the theoretical curves were estimated from data obtained under different experimental conditions with different *S*s, the model appears to possess considerable predictive power. Certainly the extreme elevation of the left segments of the Quarter groups gradients can be corrected by improvements in the apparatus. The shallowness of the central dip and the underestimation of the range of generalization may be due to the omission of factors which should enter into the application of stimulus sampling theory to generalization phenomena.

The superiority of predictions based on parameters estimated from a portion of the summated generalization data itself, suggests the possibility that much of the discrepancy between theoretical and empirical gradients in the case of a priori predictions may be attributable to sampling variation associated with the use of two different sets of *S*s in the two experiments.

SUMMARY

Two experiments were designed to test the predictive power of a mathematical model derived from the Burke-Estes (1957) stimulus sampling theory of behavior. The first, an experiment concerned with simple stimulus generalization, yielded estimates of parameters which were used to generate theoretical curves for the second, concerned with summated stimulus generalization. Procedures in the two experiments were almost identical, except that in the summated generalization experiment the training trials occurred in pairs. On training trials *S* was shown a square patch of light at one of the two training positions designated for a particular group. On test trials, if *S* believed the stimulus coincided with either of the training stimuli, he responded "same"; if he believed it differed from both stimuli, he responded "different." Treatments for the six groups differed in two ways: (*a*) there were three degrees of separation between the two training stimuli and (*b*) the range of the stimulus field differed across equal separation groups.

A priori predictions pertained to two aspects of the summated generalization gradients. First, it was expected that increasing separation between the training stimuli would lead to progressively lower response levels along intermediate portions of the gradients. Second, it was predicted that corresponding outer segments of the gradients for different separation groups would be parallel.

The two predicted relationships were generally confirmed. Empirical gradients were approximately linear in their outer segments, varying from clear bimodality in the maximum separation groups to curves with virtually horizontal ledges between the training stimuli in the minimum separation groups. Underestimation of the total range of the summated gradients and the shallowness of the curves between training stimuli constituted the systematic departures from exact predictions. However, when the parameters were estimated from the data for one of the summated generalization groups, theoretical and empirical gradients coincided rather well.

REFERENCES

BILODEAU, E. A., BROWN, J. S., & MERYMAN, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, **51**, 293–298.

BROWN, J. S., BILODEAU, E. A., & BARON, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

KALISH, H. I., & GUTTMAN, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

(Received August 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 456–460

RECOGNITION OF COMPLEX VISUAL STIMULI AS A FUNCTION OF TRAINING WITH ABSTRACTED PATTERNS[1]

MELVIN H. MARX, WILTON W. MURPHY,[2] AND AARON J. BROWNSTEIN[3]
University of Missouri

The present investigation is concerned with the general problem of recognition of complex visual stimuli when such stimuli occur repeatedly in a variety of distorted forms. The problem was orginally suggested by the task faced by the Air Force navigator, who must learn to recognize complex visual stimuli which never look exactly the same on the radar scope, even under optimal conditions. More specifically, we have been concerned with investigating the efficacy of training techniques ("briefing") for recognition of complex visual stimuli in which the pattern, or basic arrangement of the various elements, is the most stable aspect. In such stimuli the peculiarities of shape of individual elements are not stable. The hypothesis tested was that recognition of such stimuli is more effective when *S* is trained on simplified, abstracted versions of the stimulus pattern rather than the variable stimulus itself. The rationale underlying this hypothesis is that elimination of the variably shaped elements eliminates the distraction produced by this factor and thus maximizes *S*'s utilization of the more important pattern cues.

Two methods of abstraction in training were used: (*a*) presentation of a small number of straight lines representing the major direction or "flow" of the pattern, and (*b*) presentation of the pattern represented by simple geometric designs. A control briefing condition consisted of the stimulus form fully drawn in, with variably shaped elements, in the same manner as the test items were prepared. Examples of the training materials are presented in Fig. 1.

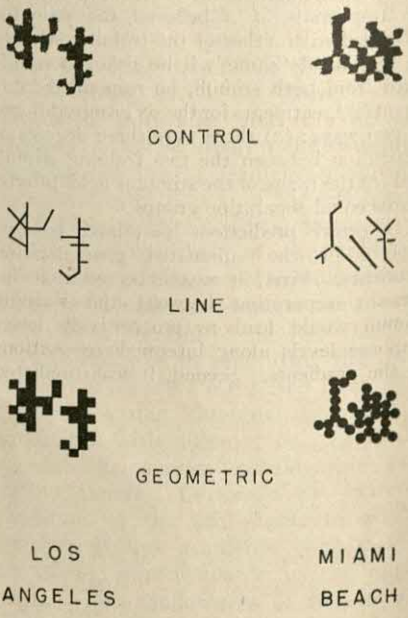


FIG. 1. Samples of training views for two patterns.

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 41(657)-11, monitored by the Operator Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

[2] Now at Wayne State University, Detroit, Michigan.

[3] Now at Florida State University, Tallahassee.

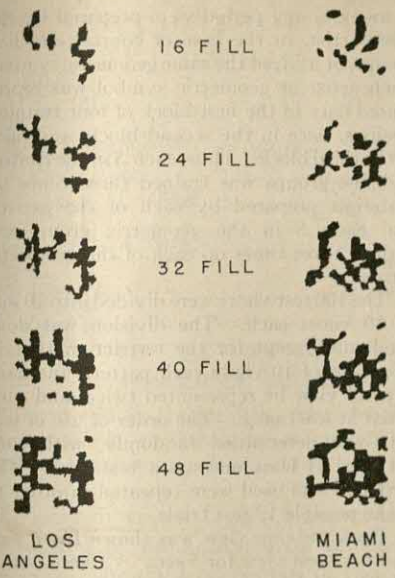


FIG. 2. Samples of test views for two patterns.

Test items consisted of two degrees of perceptual impoverishment (random removal of 25% and 50% of the original stimulus) and two degrees of visual noise (random addition of 25% and 50% irrelevant elements) as well as the complete stimulus itself. Examples of the test views are presented in Fig. 2.

METHOD

Subjects

The *S*s were 198 men and women students from three general psychology classes at the University of Missouri. Each training condition contained 66 *S*s and was represented in each class. An additional 65 *S*s were randomly discarded to equalize group *N*s to satisfy formal requirements of the analysis of variance. Statistical analyses computed with all *S*s gave the same results.

Stimulus Materials

Basic patterns.—A set of five basic patterns was prepared in the following manner. A 10 × 10 matrix was used, thus giving 100 available units. Five random and independent selections of 32 units each were made from the 100-unit matrix. The construction rules for these patterns were as follows. The 100-unit matrix was arbitrarily divided into four equal quadrants of 25 units each, the lines of division being the middle, vertical, and horizontal lines. The quadrants were numbered starting with the upper left hand quadrant as number one and proceeding in a clockwise fashion.

One of the four quadrants was picked at random to contain the one focal point, the quadrant diagonal to it being the location of the second focal point. The rules were such that the focal points must be located in these two quadrants. The particular unit in each quadrant was determined by random numbers, the one restriction being that the focal point could not be located in one of the boundary units of the quadrants. The remaining units were then added randomly with the restriction that a new unit must be contiguous with one of the units present but that contiguous boundary units could not be filled.

Control briefing.—Four "artists" were given the five basic patterns, represented by check marks indicating which units of the 10 × 10 matrix were to be filled. They were instructed to fill in the checked units by means of a ball point pen on an Azograph (A. B. Dick duplicating) master sheet, from which several copies of each preparation were made. The only rule was that each filled unit must be drawn at least 50% full, and that no unfilled unit could be drawn more than 50% full. The techniques to be used in connecting units, rounding off corners, etc., were left to the individual artist. Examples of stimuli so prepared are shown at the top of Fig. 1 and in Fig. 2.

Line training.—The same four artists were provided with overlay sheets to place on top of their control drawings. They were instructed to draw in the straight lines that they thought best represented the major parts of the pattern. They were told to keep the lines as simple as possible. Examples of stimuli so prepared are shown in Fig. 1.

Geometric training.—The pattern was depicted by means of geometric symbols, which represented the elements. In any one drawing each unit of the matrix that was to be filled was filled by the same mechanically drawn geometric symbol. Four such symbols were used—dot, empty circle, blacked circle, and blacked square—in order to provide four types of products comparable to the four different artists who prepared the other kinds of briefing stimuli. Examples of such stimuli with blacked squares and circles are shown at the bottom of Fig. 1.

Test views.—Five types of test view were prepared for each pattern. They were: (*a*) the complete 32-unit patterns; (*b*) an incomplete 16-unit view (50% impoverishment); (*c*) an incomplete 24-unit view (25% impoverishment); (*d*) the complete 32-unit pattern plus 8 added units (25% noise); (*e*) the complete 32-unit pattern plus 16 added units (50% noise).

Each of these test conditions was prepared by each artist for each pattern. Independent and random samples were taken from the population of 32 pattern units for each of the 16- and 24-unit views, and from the population of 68 noise units for each of the 8- and 16-unit noise increments. (The population of 68 noise units for any pattern consisted of those units in the matrix that did not comprise any part of the pattern.) Independent random selections were made for each of the artists.

Drawings of the 32-unit pattern itself were made independently of the control 32-unit briefing drawings, so that the 32-unit test and control briefing views were not identical even though their basic pattern was.

Photographs of the 100 test views (5 patterns, 5 types of test view, by 4 artists), as well as the briefing views, were mounted on 2 × 2 in. slides for projection by an automatic projector.

Experimental Design

Each of the three large classes was divided into three approximately equal sections, with each section receiving one of the three training conditions (control, line, or geometric). The division of a class was done according to columns of seats in an attempt to control for any motivational or ability differences that might be associated with seating arrangements. The 3 *E*s and three classrooms used for the experimental session were rotated over training conditions.

Procedure

The experimental session consisted of 12 training periods each followed by a test trial. In the presentation of the training views each pattern was identified by a city name (New York, Miami Beach, Kansas City, San Francisco, Los Angeles) which was printed plainly beneath the drawing. Test views were to be identified by the *S* by writing in the appropriate initial (NY, MB, KC, SF, LA) on the test sheet.

In each training period each pattern was shown once, in random order. All five slides of any training period were prepared by the same artist, in the case of control and line groups, or utilized the same geometric symbol. Each artist or geometric symbol was represented once in the first block of four training sessions, once in the second block, and once in the third block. Thus, each *S* in the control and line groups was trained three times on materials prepared by each of the artists, and each *S* in the geometric group was trained three times on each of the geometric forms.

The 100 test views were divided into 10 sets of 10 views each. The division was done randomly except for the restriction that in every set of 10 views each pattern and each type of view be represented twice, and each artist at least once. The order of use of test sets was determined randomly, with each set used at least once as a test trial. The first two sets used were repeated in order to make possible 12 test trials.

Each training view was shown for 7 sec., and each test view for 5 sec. A 10-sec. intertrial interval was used.

In the instructions all *S*s were told that they would be shown hypothetical visual representation of five United States cities; that their task was to learn to associate the name of each city with the appropriate pattern so that they would be able to identify it correctly under test conditions; and that after looking at the five city patterns with the names given they would be shown a number of test views, for each of which they were to record on the record sheet the initials of the city they thought was represented.

The *S*s were told that none of the test patterns would be exactly like the training patterns. It was emphasized that some of the test views would resemble the training patterns more than the others, and that some would be missing parts and others would have extra parts.

The line *S*s were told that the artists had made abstracted line drawings from the actual patterns, and geometric *S*s were told that their training patterns would be constructed of simple geometric patterns. The more fully drawn test views were also described to these *S*s.

RESULTS

Figure 3 shows the mean number of correct recognitions for the three training groups as a function of the type of test view, indicated by the number of elements, or amount of

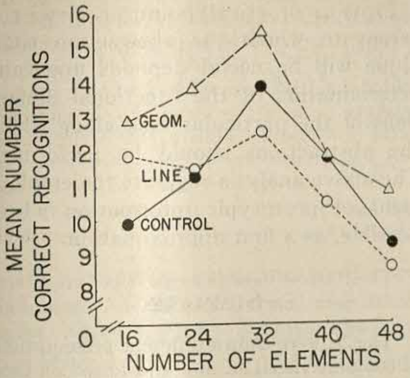


FIG. 3. Mean correct recognitions for the three training conditions as a function of number of elements (amount of fill).

fill, in each view. Table 1 shows the summary of an analysis of variance based on the data presented in Fig. 1. It can be seen that the type of training, number of elements, and the interaction of these variables were all significant sources of variance. The overall means for the three training conditions were: control = 11.29, line = 11.04, and geometric = 13.03. The geometric group mean differed reliably ($P < .01$) from the other two training conditions.

Figure 3 shows that the superiority of geometric training was consistent over the various types of test view. The differences were evaluated by t tests using the error term derived from the major analysis. The geometric group means were superior ($P = .01$) to the other two groups for all types of view except at Fill 16, where the mean differed from that for line briefing at $P < .05$, and at Fill 40, where it failed to differ reliably from the control training.

TABLE 1

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT RECOGNITIONS

Source	*df*	*MS*	*F*
Training	2	385.89	9.98*
Error	195	38.65	
Number of elements	4	509.53	87.40*
Interaction	8	40.80	7.00*
Error	789	5.83	

* $P < .01$.

With regard to the control training-line training comparison, line training was reliably better for the most impoverished test condition ($P < .01$) but control training was better for the 32- and 40-fill conditions ($P < .01$).

The overall trends indicate an asymmetry, with all groups performing better under impoverished test conditions than under noisy test conditions. This asymmetry is especially clear for the abstracted training conditions.

DISCUSSION

The results of this study show that, where *S*s are required to recognize stimuli in which the basic pattern characteristics are more stable than the elements, presentation of certain kinds of abstracted patterns results in greater recognition than presentation of the fully drawn stimuli. This conclusion is especially true for the fragmentary or impoverished type of test where both the geometric and the line abstractions were superior to the control. The superiority in this respect of the severely reduced line stimuli is particularly impressive.

These results are all the more significant in view of the fact that all test items consisted of control-type stimuli. The superiority of recognition in the experimental groups therefore occurred in spite of any positive transfer effects generated by the similarity of control training views and test views. The power of the experimental treatment is suggested by the fact that such an advantage favoring the control condition was so clearly overcome.

From a theoretical point of view, these results are of interest in that they indicate the distracting effect of overloading *S* with cues that are not stable indicators of the basic stimulus pattern (i.e., the detailed shape of the elements). The extent to which similar abstractions, whereby irrelevant or unstable cues are eliminated, may be effective in other types of learning needs to be investigated. It is important to remember that the generality of the present results is restricted to situations in which the eliminated cues are of low or questionable usefulness. The general suggestion, however, is that wherever such questionable cues can be reduced or removed acquisition efficiency may be facilitated. In further investigations the degree of stability of such cues should be directly manipulated.

Another approach to the problem is that the geometric presentation during training stood in a prototypic role to the subsequent test material in the matrix context. That is, this type of training clearly identified the matrix cells involved in the stimulus pattern. Recognition behavior would be expected more readily to generalize to variations of the prototype following learning of the prototype than following learning of similar variations of the prototype, as was the case in control training.

From a practical point of view, the extent to which the abstraction technique will be useful depends upon the determination of the functional dimensions of the particular task along which the abstractions should be performed. The above analysis suggests the employment of prototypic information where possible, as a first approximation.

SUMMARY

The *S*s were shown different presentations of complex visual stimuli and tested on their ability to identify distorted versions of the stimuli. A simple geometric technique, in which the fundamental stimulus pattern was uniformly represented by dots, circles, or squares, was found to be consistently superior as a means of training. Training which utilized only the main lines of the patterns was found to be more effective than the control (fully drawn) training stimuli for certain test items of a fragmentary sort. For test items in which the trained pattern was embedded in noise, control training was relatively more effective than line, although even here the geometric training stimuli were clearly superior.

(Received August 11, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 461–467

PARTIAL REINFORCEMENT EFFECTS (PREs) AND INVERSE PREs DETERMINED BY POSITION OF A NONREWARDED BLOCK OF RESPONSES[1]

GLEN D. JENSEN

University of New Mexico

In a runway experiment (Jensen & Cotton, 1960), a partial reinforcement effect (PRE) during Day 10 was observed in a group of rats which received 30 rewarded (R) trials on Day 1, 15 nonrewarded (NR) trials followed immediately by 15 R trials each day during Days 2–9, and 30 NR trials on Day 10. This group will be referred to as 15NR-15R. Another group (30NR-30R) which received 30 R trials on Days 1, 3, 5, 7, and 9 but 30 NR trials on Days 2, 4, 6, 8, and 10 did not show a PRE during Day 10. In the case of both of these groups, the PRE was determined in relation to a group (Control) which received 30 R trials during Days 1–9 and 30 NR trials on Day 10.

These results are consistent with the interpretation that NR trials must be patterned in such a way with R trials as to lead to the hypothesis or expectancy that reward comes after a period of nonreward and during the same session as the nonreward, if the NR trials are to produce a PRE. The 15NR-15R *S*s had eight opportunities to form such a hypothesis or expectancy prior to Day 10. The fact that these *S*s ran progressively faster during the last five NR trials of a session as Day 10 approached is consistent with the notion that they were gradually forming the expectancy that reward comes after the initial NR trials of a day are completed. Actually, another assumption is implicit: namely, that a heightened motivational state goes with this hypothesis or expectancy.

Grosslight and his associates (Grosslight & Child, 1947; Grosslight, Hall, & Murnin, 1953; Grosslight & Radlow, 1956, 1957) have marshaled considerable evidence that blocks of NR trials followed by blocks of R trials produce a PRE. However, they used very small blocks of NR trials, and their presentation does not show the gradually increasing vigor of responding during successive blocks of NR trials which seems to characterize this schedule.

EXPERIMENT I

Three of the groups of Exp. I constitute essentially a replication of the three runway groups already discussed, except that a Skinner box was used and hence single bar presses were substituted for single runway trials. The fourth group, however, underwent conditions which should produce the *inverse* of the traditional PRE if the expectancy interpretation outlined above is correct. The rats received a schedule of 20R-20NR for several days, which should result in slower pressing during the last (NR) session relative to 100% control *S*s, if it is assumed that a depressed motivational state goes with the expectancy of nonreward.

Grosslight et al. ran several groups comparable to Group 20R-20NR of

---

[1] This research was supported in full by Grant M-2834 from the United States Public Health Service.

the present study, except that they used very small blocks of NR trials. In no case did they obtain a significant inverse PRE, but one paper (Grosslight & Radlow, 1956) presents an effect which is suggestive of an inverse PRE. It was hoped that the large blocks of NR trials presently employed would produce a significant inverse PRE and also that gradually decreasing vigor of responding would occur during the NR presses on successive days of 20R-20NR scheduling.

Method

Subjects and apparatus.—The *S*s were 40 male albino rats of the Sprague-Dawley strain from 68 to 134 days in age. All *S*s were Skinner-box naive, but most of them had received 1–3 hr. of training in a runway.

A Foringer small animal test chamber was used along with Foringer programing equipment, a Gerbrands cumulative recorder, and a Davis 45-mgm. pellet dispenser.

The Gerbrands recorder was operated at the standard paper speed of 11 in. per hr. An event pen was wired through a Foringer-built contingency selector panel in such a way that every tenth bar press caused the event pen to make a horizontal mark on the paper. The distance between successive marks was used to determine the total amount of time expended during a block of 10 presses.

Procedure.—For 1 to 4 weeks before the Pretraining Day all *S*s were fed once a day with a ration of 10 gm. of finely-ground Purina laboratory chow. During the last 2 days before the Pretraining Day, all *S*s got their ration 1 hr. after the hour of the day at which they were to receive pretraining and training. The *S*s had access to water at all times except during pretraining and training sessions.

On the Pretraining Day each *S* was placed in the box for 55 min. with the apparatus set to deliver one pellet every 30 sec. until a total of 110 pellets were received. Immediately after the 55-min. session *S* was fed the remainder (5 gm.) of his usual 10-gm. ration in the form of ground chow.

Twenty-four hours after the beginning of the pretraining session each *S* was placed in the box with the bar connected to the pellet dispenser and with the automatic-delivery circuit broken. Each *S* was allowed to make 20 R presses and was then returned to his home cage where he received 9 gm. of ground chow 1 hr. after the beginning of the session. This session will be called Day 1.

The 40 rats which successfully completed Day 1 were randomly assigned to four experimental conditions except that the last 3 *S*s in a group were placed so that the four groups would have approximately equal means and variances with respect to the second 10 presses of Day 1. Group R received 40 R presses on each of Days 2–5. Group 20NR-20R received 20 NR presses followed immediately by 20 R presses during each of Days 2–5. Group 40NR-40R received 40 NR presses on Days 2 and 4, and 40 R presses on Days 3 and 5. Group 20R-20NR received reward during the first 20 presses of a day during Days 2–5 and no rewards during the last 20 presses of each of these days. All groups received a 55-min. session with no press rewarded on Day 6. All groups were 22 hr. food deprived at the beginning of each session and the 10-gm. daily ration was maintained throughout the 6 experimental days.

The time to the nearest .01 sec. taken by each *S* to make each block of 10 successive presses was determined and these values were transformed to their reciprocals divided by 10. These transformed values will henceforth be called "rate" scores.

Results and Discussion

Figure 1 shows the mean rate data for each group through the entire course of experimentation.

The representation of the long NR session given during Day 6 is complicated by the fact that *S*s in the same group made widely different numbers of presses. Arbitrarily, it was decided to represent each group for as many presses as the number made by the *S* who made the most presses. If, for example, the longest pressing *S* in Group 20NR-20R made 500 presses, and if another *S* made 200 presses, the latter *S* was assigned the arbitrary minimum rate value of .0083 for each block between 200 and 500. Note that five rate scores for each *S* in a group have been averaged in the determination of each point plotted for this 55-min. NR session. Since each *S* made at least 100 presses, the first two points are not arbitrary.

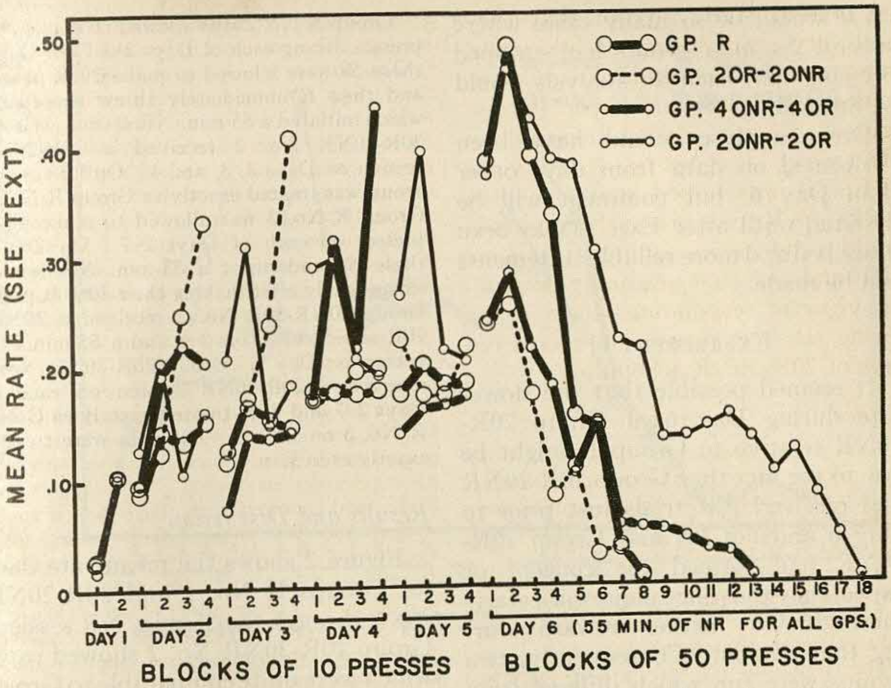


FIG. 1. The course of mean rate of pressing during all phases of experimentation.

Later data points are more or less representative of the number of "persistent" *S*s in a group as well as their rates of pressing.

The least arbitrary, but not necessarily the most sensitive test for a PRE and for an inverse PRE involves the use of the total number of presses made during the 55-min. session on Day 6. Analysis of variance on these scores yielded an *F* of 4.71 ($P < .05$; 3/36 *df*). A subsequent Tukey[2] analysis to compare individual means revealed that the 20NR-20R animals made significantly more presses than each of the other groups at the .05 level or better, but no other comparisons were significant. Thus, by this measure there is a PRE in the 20NR-20R group but there is not the predicted inverse PRE in Group 20R-20NR. There is, however, a fairly large difference in mean numbers of presses made between Groups R and 20R-20NR, namely 220.8 and 159.8, respectively.

Figure 1 shows that Groups 20R-20NR and 40NR-40R pressed at a markedly slower rate early in Day 6 relative to the other groups. An analysis of variance using the mean rate for each *S* during presses 0–100 (as determined by averaging the first 10 rate scores for each *S*) revealed that the groups differed significantly at the .03 level. A Tukey analysis revealed that Groups R and 20NR-20R showed a higher rate at the .04 level or better relative to the other two groups and no other comparison were significant. Thus both Group 20R-20NR and Group 40NR-40R showed an inverse PRE on this response measure but Group 20NR-20R did not show a PRE. If the first 300 presses were used, Group 20NR-20R probably would show a PRE, but there would be so many cases where several *S*s in a group had stopped pressing that such an analysis would be misleading.

---

[2] J. W. Tukey. The problem of multiple comparisons. Privately circulated monograph, 1953.

Other analyses could have been performed on data from days other than Day 6, but comment will be deferred until after Exp. II has been described and more reliable statements can be made.

EXPERIMENT II

It seemed possible that the slower rate during Day 6 of Group 20R-20NR relative to Group R might be due to the fact that Group 20R-20NR had received NR trials just prior to Day 6 and not because Group 20R-20NR had formed the concept or expectancy that once nonreward starts there is no further reward during that session. Consequently, two groups were run which differed from Groups R and 20R-20NR of Exp. I only in that both groups made 20 R presses immediately prior to a 55-min. NR session. Again it was predicted that the new 20R-20NR condition would produce an inverse PRE.

Three additional groups were run to determine whether a more pronounced PRE could be produced by giving seven 20NR-20R experiences rather than the four experiences given in Exp. I and to determine whether a more pronounced inverse PRE could be produced by giving six 20R-20NR experiences.

Method

Subjects and apparatus.—The *S*s were 50 male albino rats of the Sprague-Dawley strain from 69 to 131 days of age on the pretraining day. The *S*s had all received 1–9 hr. of experience in a runway and hence were quite docile.

Procedure.—The procedure was identical to Exp. I through the 20 R presses of Day 1. The first 6 *S*s of a group were assigned randomly, but the remaining 4 *S*s were placed so that the five groups of Exp. II would be comparable to each other as well as to the four groups used during Exp. I.

Group R No. 2 was allowed to make 40 R presses during each of Days 2–4. On Day 5 these *S*s were allowed to make 20 R presses and then *E* immediately threw the switch which initiated a 55-min. NR session. Group 20R-20NR No. 2 received a 20R-20NR session on Days 2, 3, and 4. On Day 5 this group was treated exactly as Group R No. 2. Group R No. 3 was allowed to make 40 R presses on each of Days 2–7. On Day 8 these *S*s underwent a 55-min. NR session immediately after making their 20th R press. Group 20NR-20R No. 2 received a 20NR-20R session on Days 2–8 and a 55-min. NR session on Day 9. Group 20R-20NR No. 3 received a 20R-20NR session on each of Days 2–7 and were treated exactly as Group R No. 3 on Day 8. The data were treated exactly as in Exp. I.

Results and Discussion

Figure 2 shows the mean rate data for Groups R No. 2 and 20R-20NR No. 2 through the 55-min. NR session. Group 20R-20NR No. 2 showed rates which were quite comparable to Group 20R-20NR of Exp. I during every phase of experimentation. Group R No. 2, however, showed markedly slower pressing during the 55-min. NR session relative to the rate exhibited by Group R of Exp. I. Analysis of variance using the total number of presses made during the 55-min. NR session yielded an *F* of less than 1.0. Thus no significant inverse PRE was demonstrated by this response measure.

Another analysis of variance was done using the mean of the first eight 10-press rates during the 55-min. NR session. One such score was obtained for each of the 20 *S*s in Groups R No. 2 and 20R-20NR No. 2. These scores represent the largest number of presses which can be used without using arbitrary scores for some *S*s. The *F* obtained is only .49, which is below the mean of the *F* distribution. Thus no inverse PRE was demonstrated by this measure either. However, the fact that Group 20R-20NR No. 2 showed rates very comparable to those of Group 20R-

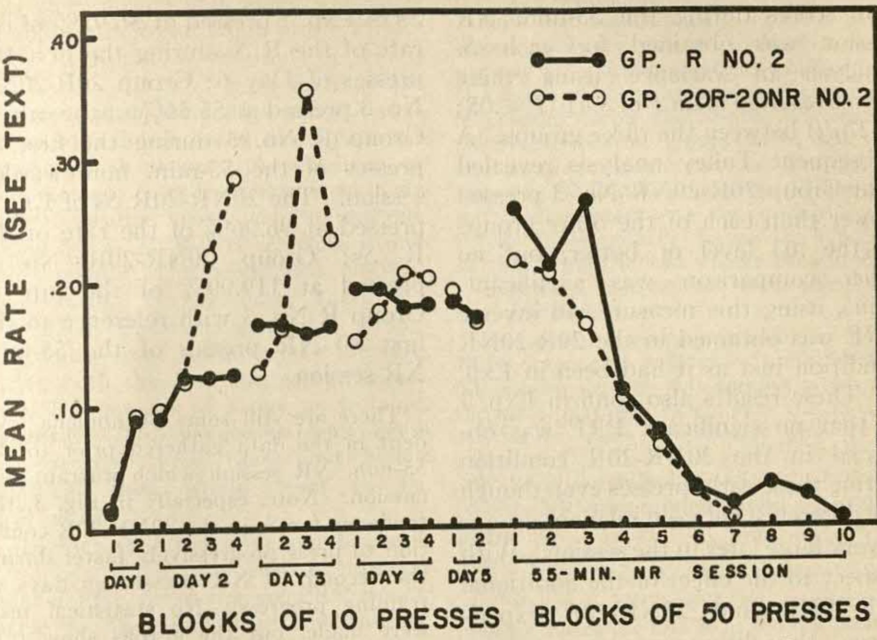


FIG. 2. Mean rate during all phases of experimentation for Groups R No. 2 and 20R-20NR No. 2.

20NR of Exp. I is encouraging, and the failure to get an inverse PRE in this part of Exp. II may yet prove to be due to the fact that Group R No. 2 was spuriously slow during nonreward.

Figure 3 shows the mean rate data for the other three groups of Exp. II. Note the close correspondence between the rates of Group 20R-20NR No. 3 and Group R No. 3 on all R blocks, especially the second R block of each day. An analysis of variance using the total number of presses made during the 55-min. NR session for each *S* yielded an *F* of 5.40 ($P < .015$; 2/27 *df*) between the three groups. A subsequent Tukey analysis revealed that Group 20NR-20R No. 2 made more presses than either of the other groups at the .02 level or better, and no other comparisons were significant. Thus, as in Exp. I, the 20NR-20R condition produced a significant PRE but the 20R-20NR condition did not produce a significant inverse PRE when total number of NR presses during 55 min. serves as the criterion. The *S*s in Group R No. 3 made an average of 240.3 presses whereas *S*s in Group 20R-20NR No. 3 made 204.4 presses on average. The percentage difference is smaller than in Exp. I, so it must be concluded that the two additional 20R-20NR experiences did not increase the inverse PRE with respect to this response measure. The *S*s in Group 20NR-20R No. 2 made an average of 375.4 presses. The percentage difference relative to Group R No. 3 is smaller than the corresponding figure from Exp. I, so it must be concluded that the extra three 20NR-20R experiences did not produce a larger PRE. These findings may mean that some value less than or equal to four 20NR-20R or 20R-20NR experiences is necessary and sufficient to produce either the PRE or the inverse PRE at the full strength, when total number of presses is the criterion.

The mean of the first eight 10-press

rate scores during the 55-min. NR session was obtained for each *S*. Analysis of variance using these scores resulted in an *F* of 3.41 ($P < .05$; 2/27 *df*) between the three groups. A subsequent Tukey analysis revealed that Group 20R-20NR No. 3 pressed slower than each of the other groups at the .03 level or better, but no other comparison was significant. Thus, using this measure, an inverse PRE was obtained in the 20R-20NR condition just as it had been in Exp. I. These results also confirm Exp. I in that no significant PRE was obtained in the 20NR-20R condition during these early presses even though the difference between relevant groups is very large later in the session. With respect to the effect of the additional 20R-20NR and 20NR-20R experiences, the differences are at least in the right direction. The 20R-20NR *S*s of Exp. I pressed at 56.97% of the rate of the R *S*s during the first 100 presses of Day 6: Group 20R-20NR No. 3 pressed at 55.56% of the rate of Group R No. 3 during the first 80 presses of the 55-min. nonrewarded session. The 20NR-20R *S*s of Exp. I pressed at 96.96% of the rate of the R *S*s: Group 20NR-20R No. 2 pressed at 119.99% of the rate of Group R No. 3 with reference to the first 80 NR presses of the 55-min. NR session.

There are still some phenomena evident in the data gathered prior to the 55-min. NR session which warrant discussion. Note, especially in Fig. 3, the tendency for *S*s in the 20NR-20R condition to press progressively faster during the second 10 NR presses as days of training progress. No statistical tests were made, but the graphs show that the actual trend is very strong and very consistent in the two 20NR-20R groups.

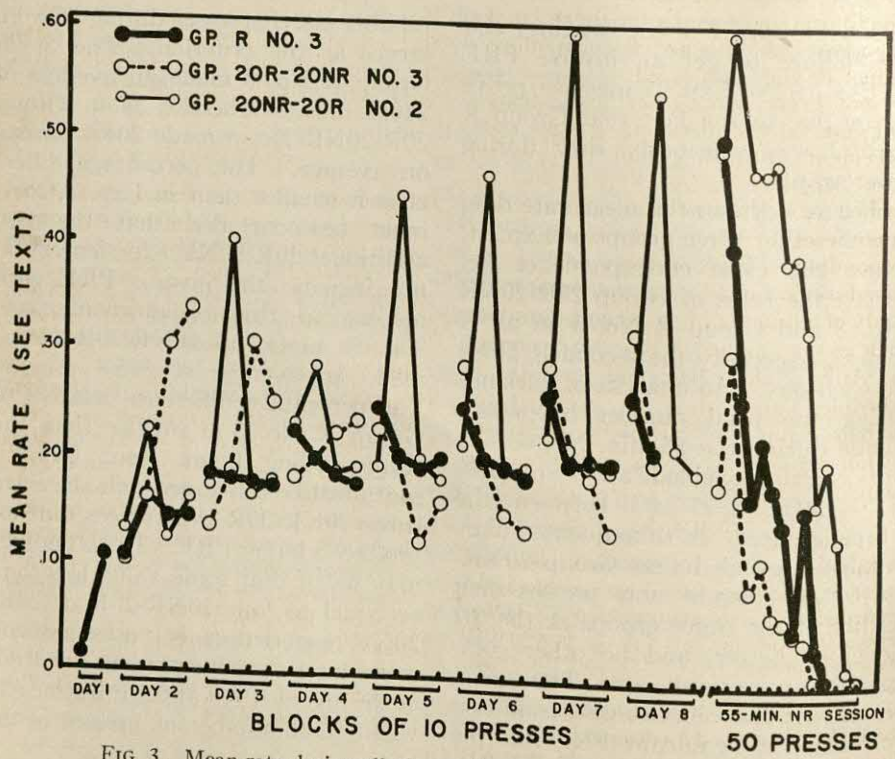


FIG. 3. Mean rate during all phases of experimentation for Groups R No. 3, 20R-20NR No. 3, and 20NR-20R No. 2.

This phenomenon is in every way comparable to the trend noted in the 15NR-15R group in the runway study by Jensen and Cotton (1960). Both findings are consistent with the notion that the rat makes hypotheses based upon past experience with respect to schedules of reinforcement. At least this is so if we further assume that the hypothesis caused by a __NR-__R schedule is one which results in a heightened motivational state. On the other hand, the 20R-20NR conditions showed a strong day-by-day trend toward *slower* pressing during both the first and second block of 10 NR presses of a day as training progresses. This phenomenon has been demonstrated three times and hence statistical tests would seem to be superfluous.

The fact that the same number and proportion of NR responses can produce either a PRE or an inverse PRE would appear to have important implications bearing upon the understanding and prediction of behavior under all conditions of partial reinforcement. If it is true that animals characteristically make hypotheses about schedules of reinforcement, as they appeared to do in the present experiments, then psychologists will not fully understand partial reinforcement until they find the general laws whereby these hypotheses are formed as well as the laws relating these hypotheses to vigor and persistence of responding. This problem can be reduced operationally to the parametric study of a number of patterning variables such as the number of R and NR presses in daily __R-__NR and __NR-__R schedules and also the number of switches from R to NR blocks of presses within sessions involving more than one switch.

SUMMARY

Nine groups of 10 rats each were run through conditions of successive blocks of rewarded (R) and nonrewarded (NR) bar presses in a Skinner box. All *S*s received 20 rewarded presses on Day 1 and the reciprocals of the time spent making the last 10 presses were used to match groups. In one condition *S*s received a pattern of 20NR-20R during each of Days 2–5 and a 55-min. NR session on Day 6. Another group received this schedule during Days 2–8 and then a 55-min. NR session during Day 9. Both groups showed a strong day-by-day trend to press faster during the last 10 NR presses of a day, and both displayed very high rates of pressing during the 55-min. NR session relative to control groups which received 40 R presses for as many days as these groups received 20NR-20R scheduling. These results are consistent with the notion that the 20NR-20R *S*s gradually formed the hypothesis that reward follows nonreward and that a heightened motivational state goes with the hypothesis.

One group received a schedule of 20R-20NR during each of Days 2–5 and a 55-min. NR session on Day 6. Another group got the same treatment except that the 55-min. NR session came immediately after 20 R presses on Day 5. A third group received the 20R-20NR schedule during Days 2–7 and a 55-min. NR session immediately after 20 R presses on Day 8. All three 20R-20NR groups showed a day-by-day trend toward slower pressing during the 20 NR presses of a day and all showed slower rates of pressing during the first 80 or 100 NR presses of the 55-min. NR session relative to the appropriate 100% control groups. These results are consistent with the interpretation that these *S*s gradually formed the hypothesis that reward does not follow nonreward within the same day and that a depressed motivational state goes with the hypothesis.

REFERENCES

GROSSLIGHT, J. H., & CHILD, I. L. Persistence as a function of previous experience of failure followed by success. *Amer. J. Psychol.*, 1947, **60**, 378–387.

GROSSLIGHT, J. H., HALL, J. F., & MURNIN, J. Patterning effect in partial reinforcement. *J. exp. Psychol.*, 1953, **46**, 103–106.

GROSSLIGHT, J. H., & RADLOW, R. Patterning effect of the nonreinforcement-reinforcement sequence in a discrimination situation. *J. comp. physiol. Psychol.*, 1956, **49**, 542–546.

GROSSLIGHT, J. H., & RADLOW, R. Patterning effect of the nonreinforcement-reinforcement sequence involving a single nonreinforced trial. *J. comp. physiol. Psychol.*, 1957, **50**, 23–25.

JENSEN, G. D., & COTTON, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, **60**, 41–49.

(Received August 15, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 468–475

THE RELATIONSHIP BETWEEN KINESTHETIC SATIATION AND INHIBITION IN ROTARY PURSUIT PERFORMANCE

RONALD S. LIPMAN[1] AND HERMAN H. SPITZ[2]

Edward R. Johnstone Training and Research Center

Several writers (Duncan, 1956; Eysenck, 1955; Köhler & Fishback, 1950) have suggested a similarity, if not an identity, between the process of satiation postulated by Köhler and Wallach (1944) to explain certain displacement effects in perception, and the process underlying some form of inhibition in learning.

Köhler and Wallach (1944) hypothesize that the act of fixating a "figure" sets certain electrochemical processes in motion in the brain, isomorphic to the contours of the figure, which, over time, create an area of impedance (satiation) that spreads in diminishing strength through the immediately surrounding brain field. As a consequence of satiation, subsequent figural currents are blocked and displaced resulting in various types of perceptual "distortion." Köhler and Wallach have employed this theory of brain processes to explain the occurrence of figural aftereffects, reversible figures, and certain illusions.

In Hullian (Hull, 1943) theory, inhibitory potential ($\dot{I}_R$) consists of two additive components: reactive inhibition (I_R) and conditioned inhibition ($_SI_R$). These constructs, extended by Kimble (1949a) and others to motor learning, have been employed to explain such seemingly diverse phenomena as reminiscence, spontaneous recovery, alternation, extinction, disinhibition, and the superiority of spaced over massed motor practice. I_R is likened to a primary negative drive state that develops whenever a motor response is made, tends to block the recurrence of the response and to dissipate spontaneously with rest. $_SI_R$, on the other hand, can be viewed as a "resting" habit in which stimuli present during I_R reduction become conditioned to I_R and when present in the learning situation tend to weaken the excitatory potential of stimulus situations otherwise evoking positive responses.

Among the writers who have suggested a similarity between satiation and inhibition, Köhler and Fishback (1950) did not identify their reference to inhibition with a particular learning theory, whereas both Duncan (1956) and Eysenck (1955) specifically referred to Hull (1943). Duncan (1956), however, rejects Hull's $_SI_R$ as a necessary construct, and limits his discussion to the similarity between Hull's I_R construct and satiation, both processes being viewed as central in locus. Eysenck (1955), for unspecified reasons, also limits his analogy to I_R. He postulates that the high satiater is the individual ". . . in whom reactive inhibition is generated quickly, in whom strong reactive inhibitions are generated, and in whom reactive inhibition is dissipated slowly . . ." (p. 97). The low satiater develops I_R slowly, in lesser amounts and dissipates it more quickly.

---

[1] The authors would like to express their grateful appreciation to Marshall P. Smith, Chairman, Psychology Department, Trenton State College, for his aid in enlisting volunteers and making space available for Exp. II. We are grateful also to Leonard S. Blackman, Belver C. Griffith, and Philip J. Siegmann of the Edward R. Johnstone Training and Research Center, Rupert Hester of the Bureau of Social Research, Department of Institutions and Agencies, State of New Jersey, and David Zeaman and Betty House of the University of Connecticut, all of whom assisted in some aspects of the present study.

[2] Now at Special Studies Unit, National Institute of Mental Health, Bethesda, Md.

The hypothesis of a similarity between satiation and inhibition has been investigated by Meier (1956) and Rechtschaffen (1958). For reasons to be discussed more fully in a later section, the results of these studies were not in agreement.

The present study reports the results of two experiments in which the rotary pursuit performance of high and low kinesthetic satiaters was compared under different work and rest schedules in an attempt to more fully test the similarity hypothesis and to identify the Hullian analogue of satiation in motor learning.

GENERAL METHOD

Subjects.—Retarded adolescents were used in Exp. I. College sophomores were employed in Exp. II.

Task order.—In both experiments *S*s were first given the kinesthetic figural aftereffect test (KAE). The *S*s with extreme KAE scores were then tested on the rotary pursuit. Each experiment was completed within a 4-mo. period.

KAE.—Since Köhler and Dinnerstein's (1947) original suggestion that kinesthetic aftereffects may be due to a satiation process, KAE has been employed widely as an index of kinesthetic satiation. Briefly, *S* is required to make equality of width judgments with the same block before and after rubbing a larger (or smaller) block. For example, when *S* rubs the larger block, the standard block feels narrower. The amount of displacement is taken as an index of the strength of the satiation process.

The KAE apparatus employed in the present study was similar to the one pictured by Klein and Krech (1952). It was mounted on two wood horses set approximately 24 in. apart. Depending on *S*'s height, *S* either stood, or sat on a stool, between the horses. A variable test (T) block, on which the position of a rider could be read to the nearest $\frac{1}{16}$ in., was situated on *S*'s right. The rider was free to move along the block from the narrow end of $\frac{3}{4}$ in. to a wide end of $2\frac{1}{4}$ in. The block was constructed so that a 1-in. change in length corresponded to a $\frac{1}{16}$-in. change in width. On *S*'s left, a rectangular T block $1\frac{1}{2}$ in. wide could be quickly replaced by a rectangular inspection (I) block $2\frac{1}{2}$ in. wide. All blocks were made of smooth hard wood, were $\frac{13}{16}$ in. thick, and had riders. The *S*s were shown the apparatus and told that they would be required to make equality of width judgments. It was demonstrated that one T block was of fixed width whereas the other T block was thin at its lower end and gradually widened. The *S* was told that he would be asked to hold the rectangular T block between the forefinger and thumb of his left hand, and the variable T block in the same manner with his right hand. By moving the right-hand rider either up or down the variable T block, he was to "find the spot on the right hand bar that feels as wide as the left hand bar." The college *S*s were then given two practice trials, while the retarded *S*s were given four practice trials (to reduce their greater variability), with feedback to the point of objective equality (POE). All trials on the KAE were alternated from random positions below and above POE. Position habits were minimized by randomly shifting the position of the variable T block before each judgment.

The normals were blindfolded before the initial practice judgments, the retardates after (for reasons of rapport). Then all *S*s made four additional judgments without feedback. The algebraic mean of these settings defined *S*'s point of subjective equality (PSE). Next, *S* firmly grasped the inspection bar and rubbed back and forth for $1\frac{1}{2}$ min. in time to a metronome set at 108 beats per min. The test bar was then quickly replaced and four more equality judgments made. The postinspection PSE was subtracted from *S*'s pre-inspection PSE to yield *S*'s mean displacement, or satiation score. These scores were measured in $\frac{1}{16}$-in. units; a + sign indicated that *S*'s postinspection PSE was smaller (in the expected direction) than his pre-inspection PSE, a − sign, the reverse.

Rotary pursuit.—A standard 60 rpm pursuit rotor and flexible stylus, with a silver head $\frac{1}{8}$ in. in diameter, were employed. Time on target scores were recorded in .001 min.

by a standard 6v. dc timer. The brass target was $\frac{3}{4}$ in. in diameter and set flush $2\frac{1}{4}$ in. from the edge of the black disc, which was $11\frac{13}{16}$ in. in diameter. The rate of rotation was checked daily and the stylus head and target were polished before testing each *S*. The *S*s were given standard instructions with *E* demonstrating how to hold the stylus and the required circular motion.

EXPERIMENT I

Method

Subjects and procedure.—Two KAE scores, separated by an interval of not more than 3 weeks, were obtained on 112 high-grade retardates (IQ above 60). The test-retest Pearson r was low (+.34) but significant ($P < .01$). The *S*s whose KAE scores were consistantly high or low on both administrations were selected for the rotary pursuit task. Group HS (high satiation) consisted of 23 males whose mean displacement scores were 3.75 mm. and 3.89 mm. Group LS (low satiation), 24 males, had mean displacement scores of −.48 mm. and −.03 mm. There was no overlap of satiation scores among the two groups. A third group—Group S (spaced)—of 18 males was selected to provide a close IQ and age match to Groups HS and LS. The mean IQs of Groups HS, LS, and S were 73, 71, and 72, respectively; the mean age, 16–6, 17–7, and 17–8.

Rotary pursuit schedule.—The massed to spaced performance schedule for Groups HS and LS consisted of 20-sec. work and 10-sec. rest for 33 trials, a 60-sec. rest, followed by 15 additional 20-sec. trials with 60-sec. intertrial rest. Group S received 48 trials of 20-sec. duration with 40-sec. intertrial rest.

Results

The results are displayed in Fig. 1. Over Trials 12–33, Group S, as anticipated, performed at a higher level than both Group LS ($P = .02$) and Group HS ($P < .001$). The Mann-Whitney U test was employed for all comparisons unless otherwise stated and the reported P values are all based on two-tailed tests. Group LS over this same block of trials performed at a significantly higher level than Group HS ($P = .02$).

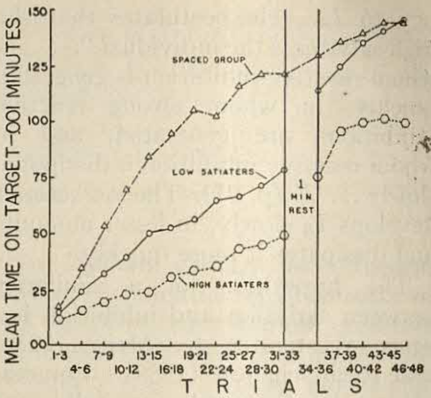


FIG. 1. Mean time on target scores for high and low satiaters under a massed to spaced schedule, and for a control group under a spaced schedule. (Trials are plotted in blocks of three. One-minute rest for massed to spaced groups only.)

The reminiscence gain (the difference between the first three spaced trials and the last three massed trials) was also greater in Group LS than in Group HS ($.10 > P > .05$). Finally, a comparison of the three groups over Trials 46–48 revealed that Groups LS and S were on target significantly more often than Group HS ($P = .02$ for both comparisons).

These results were somewhat complicated by the fact that Group LS performed at a higher level ($.10 > P > .05$) on the initial trial. In order to control for starting differences, an analysis of covariance was performed for Trials 12–33 and Trials 46–48. The corresponding *F*s of 5.27 and 7.59 (for *df* 1/44) were associated with P values below the .05 and .01 levels.

Discussion

These data support the similarity hypothesis. The high satiater develops more inhibition than the low satiater under the massed schedule and tends to dissipate less inhibition than the low satiater when switched to the spaced schedule.

The lower performance of Group HS over Trials 12–33 is difficult to interpret since it may have been attributable to a faster rate of accumulation of inhibition during work and/or a slower rate of dissipation of inhibition during the 10-sec. intertrial rest. If all inhibition is viewed as I_R, in keeping with Duncan's (1956) formulation, the reminiscence difference suggests that the dissipation of I_R during rest is a crucial factor. To test whether high satiaters also build in more I_R during work, as postulated by Eysenck, a pure massed schedule is needed.

Within Hull's two-factor inhibition theory, the most impressive finding of the present study is the asymptotic difference of Group HS relative to Group LS and to Group S. The low satiaters have not developed $_SI_R$ whereas the high satiaters are characterized by a significant amount of $_SI_R$. This, in turn, indicates that the $\dot{I}_R$ developed during massing was composed of only I_R in the low satiaters, whereas it was composed of both I_R and $_SI_R$ in the high satiaters. The presence of $_SI_R$ in Group HS but not in Group LS may have resulted from one or a combination of the following factors: (*a*) the rate at which I_R develops during work; (*b*) the level of I_R tolerated before conversion to $_SI_R$ takes place; or (*c*) the rate at which I_R dissipates during rest. In order to test possible group differences in the rate of accumulation of inhibition and to replicate earlier findings, a second experiment, employing a pure massed to spaced performance schedule, was undertaken.

A later experiment, in which high and low satiaters were administered a completely spaced schedule, is also reported here in order to establish that performance differences between satiation groups can be entirely attributed to inhibition effects. It was anticipated that performance differences between groups would not be found under the spaced schedule.

EXPERIMENT II

Method

Subjects and procedure.—In selecting *S*s for the massed to spaced schedule, the KAE was administered twice, in a single session, to 106 female college sophomores who volunteered for the experiment. Approximately half the *S*s rested 5 min. between administrations; for the remaining *S*s readministration of the KAE was about 20 min. later. This difference in test-retest time occurred because *S*s were also participating in another experiment (Spitz & Lipman, 1960). The test-retest tetrachoric r of the KAE was +.74 ($P < .001$).

The *S*s whose KAE scores were consistantly high or consistantly low on both presentations were selected for inclusion in Groups HS and LS, respectively. Three *S*s scheduled for the rotary pursuit task did not complete the experiment. Apparatus failure was responsible for the elimination of an additional *S*. Group LS consisted of 13 *S*s whose mean satiation scores were .37 mm. and .10 mm.; Group HS of 16 *S*s with mean satiation scores of 5.60 mm. and 4.46 mm. There was no overlap of satiation scores among the groups.

In view of the relatively high KAE reliability in the college population, Groups HS-S (high satiation, spaced schedule) and LS-S (low satiation, spaced schedule) were selected on the basis of a single KAE administration which was obtained on 111 female college sophomore volunteers. The *S*s with the 18 highest and lowest KAE scores were selected. Two *S*s, 1 HS-S and 1 LS-S, failed to complete the experiment. Group HS-S consisted of 17 *S*s whose mean satiation score was 6.21 mm.; the mean satiation score for the 17 *S*s of Group LS-S was .16 mm.

Rotary pursuit schedule.—Group HS and LS received 21 consecutive 20-sec. trials (7 min. pure massed) followed by a 1-min. rest and 21 additional 20-sec. spaced trials with 40-sec. intertrial rest. Continuous 20-sec. readings were taken during the pure massed trials by employing two standard 6 v. dc timers and a throw switch which *E* operated manually.

Groups HS-S and LS-S received 42 trials of 20-sec. duration and 40-sec. intertrial rest.

Results

The performance curves are shown in Fig. 2. A comparison of Groups HS and LS on Trials 1–21 (lower

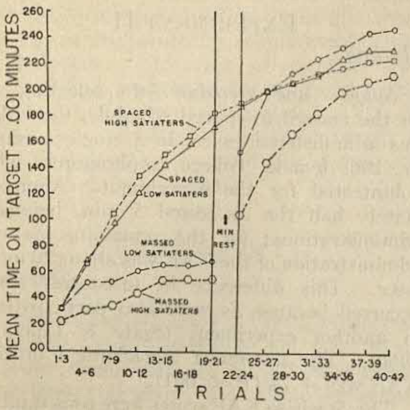


FIG. 2. Mean time on target scores for high and low satiaters under a spaced schedule, and under a massed to spaced schedule. (Trials are plotted in blocks of three. One-minute rest for massed to spaced groups only.)

curves) revealed no reliable ($P > .20$) performance differences. (This finding has since been replicated with an additional 19 low and 22 high satiaters.) The reminiscence gain (the difference between the first three spaced trials and the last three massed trials) was reliably greater in Group LS ($P < .05$) who also performed at a significantly higher level over all spaced trials ($P < .05$) and, of particular interest, over the last three spaced trials ($P < .01$). To control for the somewhat higher first trial performance of Group LS, these data were re-examined based on 10 pairs very closely matched on Trial 1 performance scores. This analysis employed the signed-rank test and confirmed the P values reported above.

An inspection of the upper curves of Fig. 2 reveals that the performance curves for the spaced practice of high and low satiaters are remarkably similar. No reliable performance difference in total time on target over all trials or over the last three trials was found. However, both spaced groups performed significantly ($P < .001$) better over Trials 1–21 than the massed practice groups of Exp. II.

A comparison of Trials 40–42 of the LS-S, HS-S, and LS groups, revealed no significant time on target differences between any of the groups. Although, as previously noted, Group LS performed at a significantly higher asymptotic level than Group HS, it was felt that the most reliable test for the presence of conditioned inhibition in Group HS would be a comparison of their asymptotic performance with that of the combined performance of the three groups (LS-S, HS-S, and LS) in which inhibition decrements could be ruled out. This comparison, over Trials 40–42, resulted in the rejection of the null hypothesis ($P = .04$).

Discussion

The similarity of the performance curves of the high and low satiaters under spaced practice strongly suggests that the performance differences found between high and low satiation groups under massed to spaced practice can safely be attributed to inhibition effects.

The failure to find a reliable performance difference between Groups LS and HS under the pure massed practice schedule, and the parallel nature of this portion of the curves, suggests that the rate of accumulation of inhibition is not an important concomitant of satiation differences. This finding plus the replication of the reminiscence and asymptotic differences between the high and low satiaters suggests, within a one-factor framework, that the rate of dissipation of I_R is the "one-factor" analogue of satiation.

When Trial 1 performance scores of high and low satiaters are pooled across groups (after randomly excluding *S*s to maintain the same number of retarded and normal *S*s in both groups) and compared by the *U* test, the resultant *P* is significant beyond the .01 level.

Thus, there is a small but statistically reliable starting difference between the high and low satiaters. In a very interesting article, Zeaman and Kaufman (1955) determined that individual differences in starting scores on a motor learning task reflected ". . . generalized conditioned inhibition ($_{s}I_{R}$), interpretable as a previously learned tendency to rest during assigned periods of prolonged work" (p. 15). In effect then, the high satiater brings somewhat more of this learned tendency to rest to the experimental situation. Moreover, since the experimental effects were significant with this learned tendency to rest statistically controlled, the finding of an asymptotic difference between Group HS and the combined scores of the other three groups (Fig. 2) indicates that the performance decrement of Group HS at asymptote can be attributed to the building in of additional $_{s}I_{R}$. Thus it must be assumed that while the amount of $\dot{I}_{R}$ developed by the high and low satiaters under the pure massed schedule (Exp. II) did not differ significantly, the composition of $\dot{I}_{R}$ did differ. That is, whereas the $\dot{I}_{R}$ developed by low satiaters consisted of only I_{R}, the $\dot{I}_{R}$ developed by high satiaters consisted of both I_{R} and $_{s}I_{R}$. In effect, then, high satiaters (in Exp. I and II) convert I_{R} to $_{s}I_{R}$ whereas low satiaters (in Exp. I andII) build in high levels of I_{R} without converting I_{R} to $_{s}I_{R}$.

It is theoretically difficult to reconcile the depressed performance of low satiaters relative to the performance of the spaced groups over the massed segments of the performance schedules with the fact that they do not develop $_{s}I_{R}$ unless it is assumed with Kimble (1949b) that I_{R} in addition to being a drive which produces resting also ". . . decreases work out-put, probably by interfering with efficient muscular co-ordination" (p. 501).

An inspection of the spaced segment of the performance curves of the high and low satiaters when switched from massed to spaced practice (Fig. 1 and 2) reveals a level but not a slope difference. That is, after massing, the performance curves of high and low satiaters, over Trials 34–48 in Exp. I and Trials 22–42 in Exp. II, are essentially parallel. The parallel nature of these curves (particularly pronounced over the earlier spaced trials) argues against different rates of dissipation of I_{R} beyond differences which have occurred during the interpolated rest.

The finding of a significantly larger reminiscence gain in the low satiaters (Exp. I and Exp. II) confirms Meier's (1956) finding of a significant negative correlation of $-.47$ between kinesthetic satiation and reminiscence following 8 min. of continuous work and a 5-min. rest on the inverted alphabet printing task. (This relatively greater reminiscence gain of low satiaters, as compared with high satiaters, cannot, of course, be generalized to relative dissipation rates of figural aftereffects.) The results of the present study and of Meier's study are inconsistent with Rechtschaffen's (1958) failure to find a statistically reliable association between *visual* satiation and an inverted alphabet printing task. However, recent research (Spitz & Lipman, 1960) indicates that visual and kinesthetic satiation are not significantly related. Therefore, Rechtschaffen's speculation that, ". . . different neural areas were being tapped in the measurement of the two phenomena" (p. 290) would explain his negative finding since the inhibition developed in the inverted alphabet task would seem to be primarily kinesthetic. In this connection, the rotary pursuit task was specifically selected for the present study on the basis of Duncan's (1957) finding that, "Reactive inhibition from massed practice on the rotor seems to be entirely the result of kinesthetic stimulation" (p. 619).

In addition to the modality problem in Rechtschaffen's study, he worked with an "inhibition score" calculated by subtracting the number of letters printed during the last 2½ min. of a 5-min. pure massed condition from a prior tally of the number of letters printed during 2½ min. of an earlier spaced condition (30-sec. work—30-sec. rest). The present findings indicate that performance correlates of satiation differences are not revealed in a pure massed or a reasonably spaced schedule.

In order to evaluate possible group differences in the rate of dissipation of I_R during the 1-min. rest in Exp. I and Exp. II it was first necessary, within Hull's framework, to gauge the amount of I_R in the high and low satiaters at the end of massing (Trials 31–33 in Exp. I, Trials 19–21 in Exp. II). In the low satiaters, their performance decrement relative to the spaced group over Trials 31–33 in Exp. I and over Trials 19–21 in Exp. II was labeled I_R. In the high satiaters (Exp. I and II), their asymptotic difference relative to the low satiaters over Trials 46–48 and 40–42, respectively, was subtracted from their performance decrement relative to the spaced group over Trials 31–33 in Exp. I and over Trials 19–21 in Exp. II, and the *remaining* decrement was labeled I_R. The reminiscence gain was then evaluated relative to the inferred amount of I_R in these groups. In both Exp. I and Exp. II the reminiscence gain of the high and low satiaters was judged to represent a very nearly equal proportion of the inferred amount of I_R present at the end of massing.

Within a one-factor framework, support was found for Eysenck's (1955) postulate that the high satiater dissipates I_R more slowly than the low satiater. However, no support was found for a different rate of I_R accumulation or differences in the amount of I_R accumulated.

Finally, then, within Hull's inhibition framework, the most parsimonious statement that can be made regarding satiation as a parameter of pursuit rotor performance is that satiation differences are reflected in the level of I_R at which conversion to $_SI_R$ takes place. The crucial difference between the high and low satiater is that the high satiater converts I_R to $_SI_R$, whereas the low satiater does not. This correspondence between the Köhler-Wallach satiation construct and Hull's $_SI_R$ seems particularly meaningful since both $_SI_R$ and satiation are conceptualized as primarily central in locus. Whether $_SI_R$ and satiation are identical processes or parallel processes characterized by similar rate constants cannot be answered with the present data.

SUMMARY

In order to test the proposed similarity between cortical satiation (Köhler & Wallach, 1944) and Hull's (1943) inhibition constructs, three groups of high and low satiaters (kinesthetic) were run on the pursuit rotor under varied conditions of work and rest. High satiaters, in Exp. I, performed at a lower level over a partially massed condition (20-sec. work—10-sec. rest), showed a smaller reminiscence gain after a subsequent 1-min. rest, and when switched to a spaced schedule (20-sec. work—60-sec. rest) performed at a lower level and showed an asymptotic decrement over the last three spaced trials. High and low satiaters in Exp. II were given a pure massed schedule (7 min.) followed by a 1-min. rest and then switched to a spaced schedule (20-sec. work—40-sec. rest). Although no significant differences were found under continuous work, the high satiaters again showed a smaller reminiscence gain and an overall and asymptotic decrement relative to the low satiaters on the spaced trials. No differences were found between high and low satiaters on a spaced schedule (20-sec.-work—40-sec. rest). A small, but statistically reliable, starting difference favoring the low satiaters was found when first trial performance scores were pooled across the groups.

These results were interpreted as indicating that the high satiater brings a somewhat stronger learned tendency to rest ($_SI_R$) to the work task and that, in addition, the high satiater converts I_R to $_SI_R$ whereas the low satiater, even after prolonged periods of continuous work, does not convert I_R into $_SI_R$.

REFERENCES

Duncan, C. P. On the similarity between reactive inhibition and neural satiation. *Amer. J. Psychol.*, 1956, **69**, 227–235.

Duncan, C. P. Visual and kinesthetic components of reactive inhibition. *Amer. J. Psychol.*, 1957, **70**, 616–619.

Eysenck, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23. (a)

Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510. (b)

Klein, G. S., & Krech, D. Cortical conductivity in the brain injured. *J. Pers.*, 1952, **21**, 118–148.

Köhler, W., & Dinnerstein, D. Figural after-effects in kinesthesis. In *Miscellanea psychologica Albert Michotte.* Paris: Librairie Philosophique, 1947. Pp. 196–220.

Köhler, W., & Fishback, J. The destruction of the Müller-Lyer illusion in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950, **40**, 398–410.

Köhler, W., & Wallach, H. Figural aftereffects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

Meier, J. M. Interrelationships among MMPI variables, kinesthetic figural aftereffect, and reminiscence in motor learning. Unpublished doctoral dissertation, University of Wisconsin, 1956. (Mic. 18–426, University Microfilms, Ann Arbor, Mich.)

Rechtschaffen, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

Spitz, H. H., & Lipman, R. S. The reliability and intercorrelation of individual differences on visual and kinesthetic figural aftereffects. *Percept. mot. Skills*, 1960, **10**, 159–166.

Zeaman, D., & Kaufman, H. Individual differences and theory in a motor learning task. *Psychol. Monogr.*, 1955, **69**,(6, Whole No. 391).

(Received August 15, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 476–483

CONFLICT AND THE ORIENTATION REACTION

D. E. BERLYNE[1]

Boston University

Among the many factors that have been shown to influence the strength and direction of exploratory behavior, one particular group of closely related stimulus properties predominates (Berlyne, 1960a). These properties may be called "collative" properties, since they depend on the collation of information from different sources. They include degree of novelty, surprisingness, complexity, and power to induce uncertainty.

Meanwhile, other investigations have shown at least some forms of exploratory behavior to be part of a pervasive complex of psychophysiological processes which mobilize the organism's equipment for gathering information through sensory channels and for taking drastic action. This complex, known as the "orientation reaction" (Berlyne, 1960a, Ch. 4; Sokolov, 1958), involves both transitory jumps and more lasting upward drifts in the level of arousal.

The experiments to be reported were concerned with the third side of the triangle, namely the relations between collative variables and arousal. There is no lack of evidence that the level of arousal and the magnitude of the orientation reaction are affected by collative variables in general. But several logically distinct collative variables have generally been confounded in the relevant studies, just as they are often correlated with one another in naturally occurring situations.

In particular, the present experiments were meant to test the hypothesis that the motivational effects common to collative variables are due to conflict. The bearing of this hypothesis on exploratory and epistemic behavior has been examined from a number of angles by the writer (Berlyne, 1954, 1957b, 1960a, 1960b). A related view has been expressed by Polezhaev (1958, 1959a, 1959b, 1960), who finds that the orientation reaction appears whenever antagonistic activities are simultaneously called forth in the dog.

The hypothesis implies that the intensity of the orientation reaction will increase with degree of conflict. Experiments directed at the effects of conflict on indices of arousal have been scanty. There have been studies of disturbances in behavior due to conflict (e.g., Hovland & Sears, 1938), but these have generally confounded degree of conflict with novelty, surprisingness, and stimulus intensity. In order to test our hypothesis regarding the role of conflict, we must devise a situation in which the degree of conflict can be varied without, as far as possible, being correlated with other variables that might affect arousal. We must also sample arbitrary forms of conflict, since most previous investigators have concentrated on special cases, e.g., conflicts between affective evaluations (e.g., Lanier, 1941) or between vital activities (e.g., Polezhaev, 1958, 1959a, 1959b, 1960), and these may well have special properties apart from those possessed by conflict in general.

---

[1] The author is greatly indebted to William E. Carr, who assisted at all stages of this series of experiments, and to Paul Weisberg, who assisted in the analysis of the data.

The experiments were carried out at the National Institute of Mental Health.

In the present experiments, the dependent variable was the galvanic skin response (GSR), recorded by the Fels Dermohmmeter. The GSR is recognized as one of the vegetative components of the orientation reaction, and skin conductance is widely accepted as an index of level of arousal. The GSR results from direct stimulation of points throughout the brain-stem reticular formation in the flaxedilized cat (Bloch & Bonvallet, 1960), and the reticular formation is apparently the part of the nervous system that has most to do with the orientation reaction (Berlyne, 1960a, Ch. 4).

Finger volume was also recorded by means of a plastic capsule enclosing the middle finger of the left hand and connected through a Statham pressure transducer to a Sanborn polygraph recorder. But since most of the fluctuations in the baseline did not appear to correspond to presentations of stimuli, this measure will not be discussed in what follows.

EXPERIMENT I

In Exp. I, the situations selected as high-conflict and low-conflict conditions were, respectively, the free-choice and forced-choice situations that had previously proved serviceable in a series of experiments on conflict and reaction time (Berlyne, 1957a). The free-choice situation is one in which two visual stimuli corresponding to different key pressing responses appear at once, and *S* has to respond to either stimulus but not both. In the forced-choice situation, there is simply one stimulus corresponding to one response, so that *S* has no freedom of choice as far as his response is concerned. We speak of "high-conflict" and "low-conflict" conditions, rather than of "conflict" and "nonconflict" conditions, because, even in a forced-choice situation, there will presumably be some degree of conflict between the correct response and other, erroneous response tendencies resulting from stimulus generalization.

Method

If the usual procedure of having a short-lasting stimulus and instructing *S* to respond to its onset had been followed, there would generally have been one GSR per trial, and this GSR would have reflected the joint effects of stimulus onset, stimulus termination, and the motor process (including proprioceptive stimulation consequent on the performance of the response) (Terekhova, 1958). It was therefore decided to have a stimulus pattern lasting 10 sec. and to instruct *S*s to respond to its *termination.* There were thus generally two GSRs per trial. The second of these must have reflected the joint effects of stimulus termination and response, whereas the first GSR was able to represent the effect of being exposed to stimulus elements associated with incompatible responses, without being contaminated by any motor process. This procedure was called for, not only as a means of increasing the sensitivity of the measure of primary interest, but also as a control. High-conflict trials might possibly have produced more vigorous responses, resulting in GSRs of higher amplitude. If the effects on skin conductance of stimulus onset and response had been allowed to coincide, an invalid conclusion that high-conflict stimulus patterns produce larger GSRs might thus have been favored.

Apparatus.—Apart from the Dermohmmeter and plethysmographic equipment, the apparatus consisted of a stimulus panel, a response unit, and a programer. The stimulus panel bore eight GE 44 6-v. bulbs covered with yellow jewels, two at each of the four corners of a diamond. The panel was painted gray, except for a white fixation spot in the center of the diamond. The response unit included a key that could be moved vertically or horizontally in four directions, corresponding to the corners of the diamond.

The programer operated by means of a punched tape and was set to illuminate combinations of the eight lights in a predetermined sequence, each combination lasting 10 sec. and separated by an 11-sec. interval from the next.

Controls.—For low-conflict trials, two lights at one corner of the diamond were illuminated (hereafter referred to as an *adjacent* pattern), whereas, for high-conflict trials, two lights at different corners were illuminated (hereafter referred to as a *dispersed* pattern).

Degree of conflict was thus isolated from stimulus intensity, since two lights were on during trials of both types. The adjacent patterns comprised one inner light and one outer light. It was thus necessary to control for distance from fixation point by arranging for half of the dispersed patterns to comprise two inner lights and the other half to comprise two outer lights.

High-conflict and low-conflict situations did not differ in novelty or surprisingness, since *S*s were apprised beforehand of the nature of the two kinds of trials and informed that they would be intermingled. Moreover, two sequences were used with equal numbers of *S*s, each having adjacent patterns where the other had dispersed patterns and vice versa.

There remains the possibility that different magnitudes of GSR may be produced by adjacent and dispersed stimulus patterns as such, regardless of the degree of conflict between overt responses that they occasioned. They may, for example, be regarded as differing in complexity. To test for this possibility, *S*s were put through a preliminary sequence of adjacent and dispersed patterns in connection with which no overt response was required or even mentioned.

Subjects.—There were 24 *S*s, 12 men and 12 women. The men were 10 normal-control volunteer inpatients in their late teens or early twenties, one library assistant, and one research assistant in social psychology. The women were eight psychiatric nurses, one typist, one administrative assistant, one librarian, and one research sociologist.

Procedure.—The two sequences, X1 and X2, each comprised 12 adjacent patterns and 12 dispersed patterns in a random order. The 4 possible adjacent patterns were used three times each, and the 12 possible dispersed patterns of two inner or two outer lights once each. Half of the *S*s in each sex group had X1 for Phase I and X2 for Phase II, and the remainder had the sequences in the opposite order.

The *S* was seated in a dimly lit room in front of the table, facing the fixation spot. All recording and programing equipment was located in an adjoining room, in which the 2 *E*s stationed themselves during the presentation of stimuli. The capsule for measuring finger volume was attached to the middle finger of *S*'s left hand. The GSR electrodes were attached to the soles of the feet, since the right hand was to be used for key pressing, and the record would have been loaded with movement artifacts if the palmar GSR had been measured. There was an interval of 15 min. for skin hydration. Then *S* underwent the two experimental phases.

In Phase I, *S* was told that lights would go on and off. Sometimes there would be two lights at the same corner, and at other times there would be two lights at different corners. He was to do nothing but sit back with his eyes fixed on the fixation spot, trying not to look at the lights directly. He was also to keep his feet and left hand as still as possible, moving them, if necessary, only while no lights were on. The sequence of stimulus patterns was then presented.

In Phase II, *S* was shown the response unit, and it was explained that the key could be moved in any of four directions, corresponding to the four corners of the diamond. There would be a sequence of lights like the one he had just experienced. He was to keep the key grasped between the thumb and first two fingers of the right hand and to move the key as soon as the lights went off. When there were two lights at the same corner, he was to move the key in the direction indicated by the lights, as soon as they disappeared. When there were two lights at different corners, he was to move the key in either, but not both, of the directions indicated by the lights, as soon as the lights disappeared. Once again, he was to keep his eyes on the fixation spot and keep his feet and left hand still. The second sequence was then presented.

Results

For each *S*, three sets of GSRs were examined, namely, (*a*) stimulus-onset GSRs for Phase I, (*b*) stimulus-onset GSRs for Phase II, and (*c*) stimulus-termination-and-response GSRs for Phase II. A GSR was recognized if there was a downward deflection in the skin-resistance tracing during the action of the stimulus pattern in the case of *a* and *b* or during the interval between two successive patterns in the case of *c*. The point at which the first competent deflection occurred was treated as the initial point, and the point of minimum resistance reached by any swing starting with

TABLE 1

Experiment I: Transformed GSR Scores

Phase	Adjacent Patterns (Low Conflict in Phase II)		Dispersed Patterns (High Conflict in Phase II)	
	Mean	*SD*	Mean	*SD*
I	0.806	0.812	0.768	0.748
II: S onset	0.924	0.711	1.120	0.724
II: S termination and R	1.557	0.748	1.613	0.654

a competent deflection was treated as a peak-point.

The resistance values read from the tracing were subjected to the log-conductance-change transformation recommended by Haggard (1949) for data intended for analysis of variance. The actual score used was $\log_{10}[10^9(1/R_P - 1/R_I) + 1]$ where R_P and R_I are the resistances, in ohms, at the peak-point and the initial point, respectively.

For each *S*, a mean adjacent-pattern (low-conflict in Phase II) score and a mean dispersed-pattern (high-conflict in Phase II) score were comcomputed for Phase I stimulus-onset, Phase II stimulus-onset, and Phase II stimulus-termination-and-response GSRs. The means are exhibited in Table 1 and the results of analyses of variance in Table 2.

The comparison of most interest is that between dispersed-pattern (high-conflict) trials and adjacent-pattern (low-conflict) trials for Phase II stimulus-onset GSRs. It will be seen that the mean for the former is higher ($P < .01$), as would be predicted from the hypothesis that the intensity of the orientation reaction increases with conflict. The corresponding comparison for Phase I, in which there were no overt responses, produces a nonsignificant difference in the opposite direction. The difference in Phase II cannot therefore be attributed to intrinsic properties of the stimulus patterns, and we may regard the hypothesis that the magnitude of the orientation reaction increases with degree of

TABLE 2

Analysis of Variance for Exp. I

Source	*df*	Phase I Stimulus Onset		Phase II Stimulus Onset		Phase II Stimulus Termination and Response	
		MS	*F*	*MS*	*F*	*MS*	*F*
Between *S*s							
Order of sequences (O)	1	1.631	1.51	1.488	1.69	1.932	2.28
Sex groups (S)	1	5.437	5.05*	2.886	3.29	0.700	0.83
O × S	1	0.211	0.20	1.115	1.27	2.235	2.75
Residual	20	1.077	76.93***	0.878	16.88***	0.846	13.02***
Total between *S*s	23						
Within *S*s							
Adjacent/Dispersed patterns (A/D)	1	0.017	1.21	0.461	8.87**	0.037	0.57
O × A/D	1	0.177	8.36**	0.566	10.88**	0.581	8.94**
S × A/D	1	0.049	3.50	0.018	0.35	0.015	0.23
O × S × A/D	1	0.014	1.00	0.015	0.29	0.011	0.17
Residual	20	0.014		0.052		0.065	
Total within *S*s	24						
Grand total	47						

* $P < .05$.
** $P < .01$.
*** $P < .001$.

conflict as confirmed. It may be noted that the stimulus-termination-and-response GSRs in Phase II are higher after high-conflict than after low-conflict trials, but the difference is not significant.

In all three analyses of variance, the interaction between adjacent vs. dispersed patterns and order of sequences is significant. This simply reflects the fact that half of the *S*s had Sequence X1 and the other half Sequence X2 in each phase. Sequence X1 began with two dispersed patterns and X2 with two adjacent patterns, and the first two trials always gave much bigger GSRs than subsequent trials. The sex difference is significant for Phase I but not for Phase II. The male *S*s had a higher mean (1.123 as compared with 0.450) in Phase I than female *S*s; however, it will be recalled that the two sex groups differed in other respects.

The overall mean for Phase II stimulus-termination-and-response GSRs exceeded that for Phase II stimulus-onset GSRs ($Ms = 3.170$, 2.044; $t = 4.02$, $df = 23$, $P < .001$). And the mean for Phase II stimulus-onset GSRs exceeded that for Phase I stimulus-onset GSRs ($Ms = 2.044$, 1.574; $t = 2.39$, $df = 23$, $P < .05$). The latter result confirms what Russian experimenters have repeatedly found (e.g., Sokolov 1958, Ch. 3), viz., that the orientation reaction is more pronounced when stimuli have to be discriminated for the guidance of an overt response or, as Russian writers put it, when the stimuli are "signal stimuli."

EXPERIMENT II

Experiment II studied different degrees of a more complex kind of conflict, namely response uncertainty in a word association situation. The magnitude of the GSR was once again used as an index of the strength of the orientation reaction.

Laffal (1955) administered 100 stimulus words to 80 *S*s calling for free-association responses. He tabulated the alternative responses given to each stimulus word together with their relative frequencies and calculated, for each stimulus word, a response-uncertainty (*entropy*, H) value according to the usual information-theory formula: $H = -\Sigma_i p_i \log_2 p_i$.

He argued that, since verbal associations are primarily culturally determined, each *S* can be assumed to have possessed a hierarchy of responses to each word, approximating the set of responses given by the population as a whole, with relative strengths approximating their relative inter-*S* frequencies. The inter-*S* response uncertainty thus provides an estimate of the mean intra-*S* response uncertainty. It was found that stimulus words with higher response uncertainties occasioned more "response faults," e.g., failure to recall the response when the stimulus word was presented a second time, exceptionally high latency, misunderstanding.

A stimulus word with which a number of competing verbal responses are associated is, of course, a conflict-inducing stimulus, and, in such circumstances, the information-theory measure of uncertainty can be regarded as a measure of the degree of conflict (Berlyne, 1957b, 1960a). It should be pointed out that it is here a matter of an observer's uncertainty about the response that *S* will make. This must not be confused with the collative stimulus property we have called "power to induce uncertainty," which refers to the generation of uncertainty inside *S* with respect to what will follow, although conflict is involved in both cases and the same information-theoretic measure is applicable.

Method

Material.—Six high-uncertainty and six low-uncertainty words were selected from Laffal's list. All of the words at the upper end of Laffal's list that had threatening connotations and also the word "suck" with its possible oral-erotic significance were omitted. Otherwise, the words selected (shown in Table 3) were the six with the highest response uncertainty and the six with the lowest response uncertainty cited by Laffal. The high-uncertainty words had a mean H of 4.7264 ($SD = 0.32$) and the low-uncertainty words a mean H of 1.2294 ($SD = 1.17$), the means being significantly different at the .001 level ($t = 15.75$, 10 df). It will be noted that there were "emotional" words in both categories.

The 12 words were arranged in a randomized sequence with the restriction that each successive pair of words had a high-uncertainty and a low-uncertainty member. The sequence began with a low-uncertainty word and ended with a high-uncertainty word.

Subjects.—There were 13 men and 7 women, all of whom participated in Exp. II either immediately after Exp. I or after pilot work for Exp. I. Ten men heard the stimulus words in the original sequence mentioned above. The remaining *S*s heard them in the reverse order.

Procedure.—The *S*s were instructed to close their eyes and listen while 1 *E*, sitting in the same room, read out a list of words. They were to respond in each case with the first word that came to mind. The words were read at the rate of one word every 21 sec.

Results

As Table 3 shows, the mean transformed GSR score for the high-uncertainty words, 2.508, was higher ($P < .01$, see Table 4) than the score for the low-uncertainty words, 2.332. The conflict hypothesis may thus be considered to have passed a further test.

TABLE 3

STIMULUS WORDS IN EXP. II

Word	Position in Sequence		H (Bits)	Mean GSR Score
	Original	Reversed		
High-Uncertainty Words				
SPIT	12	1	4.9380	2.776
TANGLE	6	7	4.8768	2.344
SOOTHE	10	3	4.6808	2.539
SNEAK	8	5	4.6448	2.680
WRINKLE	2	11	4.6208	2.495
DUMP	3	10	4.5975	2.217
Means			4.7264	2.508
Low-Uncertainty Words				
DWARF	5	8	1.9952	2.340
STAMMER	9	4	1.5355	2.374
STUMBLE	11	2	1.4979	1.916
SHOVE	7	6	0.9337	2.278
GARAGE	1	12	0.8088	2.776
RIP	4	9	0.6238	2.312
Means			1.2294	2.332

TABLE 4

ANALYSIS OF VARIANCE FOR EXP. II

Source	*df*	*MS*	*F*
Between *S*s			
Stimulus order (O)	1	0.675	0.03
Residual	18	26.517	28.85**
Total between *S*s	19		
Within *S*s			
Response uncertainty (*H*)	1	9.793	10.66**
O × *H*	1	0.204	0.22
Residual	18	0.919	
Total within *S*s	20		
Grand total	39		

** $P < .01$.

EXPERIMENT III

Numerous experiments have established that the level of arousal and the magnitude of the orientation reaction are affected by stimulus novelty unconfounded with surprise. Separate components of the orientation reaction have been shown to weaken and finally extinguish altogether when a monotonous stimulus is repeated (Berlyne, 1960a). Sharpless and Jasper (1956) demonstrated similar habituation when they recorded the electrical activity of the reticular formation directly. Whether or not the first stimulus in such repetitious sequences is surprising or not for a human *S* depends on how far he has been told beforehand what to expect. But the stimuli will certainly no longer surprise after a few presentations, and each succeeding stimulus will differ from previous ones only in being less novel.

Similarly, many investigators have shown that the sudden unexpected introduction of an alteration in experimental conditions will intensify or disinhibit the orientation reaction. But we need to ascertain whether arousal is affected by surprisingness unconfounded with novelty. This was the aim of Experiment III.

The conflict hypothesis interprets surprise as a special form of conflict, on the grounds that an expectation will bring out one kind of response and a stimulus disagreeing with the expectation will bring out another kind of response that is incompatible with the first. The fact that conflict independent of surprise has been found to influence the orientation reaction in Exp. I and II favors this interpretation.

Stimuli that were surprising without being novel were contrived in the following manner. The apparatus was the same as that used in Exp. I. The expectations to be violated were built up both by prior verbal description and by repetition of a temporal pattern. Let us call the two lights at the upper corner of the diamond "A" and the two lights at the right-hand corner "B." For one sequence, U1, *S* was told that A and B would light up in turn. The sequence actually consisted of ABABABABABABABA**A**AB*A*B. At Position 16 the **A** appeared in place of a B and is thus designated the *surprising* stimulus. The *A* in Position 19, which agreed with the expectation of alternation, was designated the *comparison* stimulus. As for the other sequence, U2, *S* was told beforehand that it would consist of B and A in alternation, but it actually consisted of BABABABABABABAB*A*BA**A**A. Here, the **A** in Position 19 was the surprising stimulus and the *A* in Position 16 was the comparison stimulus. The stimuli represented by **A**'s were surprising but not novel, since the sequence had already contained several A's and the last stimulus encountered was, in fact, an A.

All *S*s had both sequences, one with an overt key pressing response required and one without the overt response, order being counterbalanced. GSR data were recorded and analyzed as in the other two experiments.

Subjects.—There were 5 men and 7 women *S*s, all of whom participated in Exp. III immediately after Exp. I or after pilot work for Exp. I. No. *S* participated in both Exp. II and III. All *S*s went through two phases, 3 *S*s having each of the following orders of treatments (R = "overt response"): U1R-U2NoR, U2R-U1NoR, U1NoR-U2R, U2NoR-U1R.

Procedure.—Stimuli were programed to be on for 3 sec. with 9-sec. intervals between them. Before sequences with overt responses, *S*s were instructed to respond to the onset of each stimulus pattern by moving the key in the corresponding direction. Before sequences without overt responses, they were told to sit back and do nothing. In both cases, they were to keep their eyes on the fixation spot and to keep their feet still.

After the first phase, *S*s were told that one of the stimuli had appeared out of place because of a hitch in the working of the equipment but it was hoped that such a hitch would not occur during the next phase.

Results

Mean transformed GSR scores are displayed in Table 5. Analysis of variance (see Table 6) shows the mean for surprising stimuli to be significantly higher ($P < .001$) than the mean for comparison stimuli. Particularly interesting, however, is the fact that the interaction between the surprising vs. comparison stimuli and presence vs. absence of overt response did not even approach significance. What an interaction means depends, of course, on what kind of transformation is applied to the data. With our logarithmic transformation, the lack of a significant interaction means that we have no reason to believe that the *proportion* by which surprise enhances the rise in conductance is affected by the presence or absence of an overt response.

TABLE 5

MEAN TRANSFORMED GSR SCORES: EXP. III

Condition	Surprising Stimuli		Comparison Stimuli	
	Mean	*SD*	Mean	*SD*
With overt R	1.918	1.056	1.132	1.026
Without overt R	0.739	0.870	0.236	0.624

TABLE 6

ANALYSIS OF VARIANCE FOR EXP. III

Source	*df*	*MS*	*F*
Within *S*s: Between Phases			
Presence/Absence of overt response (R)	1	12.909	14.72**
Sequences (S)	1	1.832	2.09
Phases (P)	1	0.186	0.21
Residual	9	0.877	
Total	12		
Within *S*s: Within Phases			
Surprising/Comparison stimuli (St)	1	4.987	24.33***
R × St	1	0.240	1.17
S × St	1	0.025	0.12
P × St	1	0.064	0.31
St × *S*s	11	0.704	3.43*
Residual	9	0.205	
Total	24		

* $P < .05$.
** $P < .01$.
*** $P < .001$.

Some disturbance is only to be expected when *S* is set to receive a stimulus calling for one sort of overt action and receives instead a stimulus calling for another sort of action. But it is instructive, especially in view of its bearing on intrinsic or "disinterested" exploratory activity, to find that a surprising event is no less arousing when it does not have any implications for *S*'s overt behavior. Several *S*s actually reported that they had not been paying attention during the phase without overt responses and had not noticed that one of the stimuli appeared out of turn.

Once again, we see that GSRs were significantly greater in amplitude when overt responses were performed than when they were not. This confirms further that "signal stimuli" produce greater orientation reactions than stimuli without signal value.

SUMMARY

Three experiments were carried out with the dual aim of studying the influence of collative stimulus properties on the amplitude of the orientation reaction and testing the hypothesis that conflict underlies the motivational effects of collative variables. The GSR was used as an index of the orientation reaction.

In Exp. I, using forced-choice and free-choice reactions, GSR amplitude increased with degree of conflict, unconfounded with novelty, surprisingness, intensity, and distance from fixation point, of stimuli. A preliminary phase with no overt responses showed that the effect was not due to inherent properties of the stimulus patterns, e.g., differences in complexity. In Exp. II, using word association, GSR amplitude increased with response uncertainty, regarded as a measure of degree of conflict. In Exp. III, stimuli that were surprising without being novel, surprise being interpreted as a form of conflict, likewise produced more intense GSRs than others.

REFERENCES

Berlyne, D. E. A theory of human curiosity. *Brit. J. Psychol.*, 1954, **45**, 180–191.

Berlyne, D. E. Conflict and choice time. *Brit. J. Psychol.*, 1957, **48**, 106–118. (a)

Berlyne, D. E. Uncertainty and conflict: A point of contact between information-theory and behavior-theory concepts. *Psychol. Rev.*, 1957, **64**, 329–339. (b)

Berlyne, D. E. *Conflict, arousal and curiosity.* New York: McGraw-Hill, 1960. (a)

Berlyne, D. E. Les équivalences psychologiques et les notions quantitatives. In D. E. Berlyne & J. Piaget, *Théorie du comportement et opérations.* (*Etudes d'épistémologie génétique,* No. 12) Paris: Presses Univer. France, 1960. (b)

Bloch, V., & Bonvallet, M. Le déclenchement des réponses électrodermales à partir du système réticulaire facilitateur. *J. Physiol., Paris,* 1960, **52**, 25–26.

Haggard, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39**, 378–392.

Hovland, C. I., & Sears, R. R. Experiments on motor conflict: I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, **23**, 477–493.

Laffal, J. Response faults in word association as a function of response entropy. *J. abnorm. soc. Psychol.*, 1955, **50**, 265–270.

Lanier, L. H. An experimental study of "affective conflict." *J. Psychol.*, 1941, **11**, 199–217.

Polezhaev, E. F. [The conditioned postural reflex as a special reaction during the interval between presentations of stimuli.] *Dok. Akad. Nauk SSSR,* 1958, **123**, 204–207.

Polezhaev, E. F. [On cortical co-ordination during the transition between sleep and wakefulness.] *Dok. Akad. Nauk SSSR,* 1959, **126**, 1149–1152. (a)

Polezhaev, E. F. [Novelty as a stimulus for special reactions.] *Byull. eksp. Biol. Med.*, 1959, **2**, 9–14. (b)

Polezhaev, E. F. [Phase changes in EEG as indicators of the formation of cortical co-ordinations.] *Fiziol. Zh. SSSR,* 1960, **46**, 26–36.

Sharpless, S., & Jasper, H. H. Habituation of the arousal reaction. *Brain,* 1956, **79**, 655–680.

Sokolov, E. N. *Vospriiate i uslovny refleks.* [Perception and the conditioned reflex.] Moscow: Moscow Univer. Press, 1958.

Terekhova, O. P. [The orientation reflex in conditioned reactions to complex stimuli.] *Vop. Psikhol.*, 1958, 4(1), 87–96.

(Received August 19, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 484–495

SOME RELATIONS BETWEEN THE INTENSIVE PROPERTIES OF THE CONSUMMATORY RESPONSE AND REINFORCEMENT[1]

GEORGE COLLIER, FREDERICK A. KNARR, AND MELVIN H. MARX

University of Missouri

The present experiment is part of a series (Collier & Myers, 1961; Knarr & Collier, 1960) by which it is hoped to partial out the contributions of the various intensive properties of the consummatory response to reinforcement magnitude. By using a drinking tube in the runway as a magazine, various aspects of the evoking stimuli and the consummatory response can be manipulated. For example, by using solutions varying in quality or concentration the evoking stimuli can be varied. By varying the tube size and by removing the animal from the goalbox after various durations of drinking, numbers of licks, or volumes drunk, these factors can be varied and their effect evaluated.

In the present study the duration of the contact with the magazine was held constant and made sufficiently short so that licking was continuous. Sucrose solutions of 4% and 32% were used. The concentration variable is known to affect the level of instrumental performance, and it is presumed to affect the associated consummatory performance. The present study provides an opportunity to examine their conjoint variation in a situation where postingestive and nutritive factors play a minimal role because of the brief magazine contact and small volumes consumed. Two additional factors which have been hypothesized as determinants of the magnitude of consummatory responding, delay in reinforcement and shifts in reinforcement, have been varied. The manipulation of these variables provides the opportunity of observing the conjoint changes in consummatory and instrumental responses as a function of changes in the character of reinforcement.

METHOD

Subjects.—The *S*s were 26 female, 150-day-old albino rats from the University of Missouri colony. Two *S*s were discarded for failure on three consecutive trials to traverse the runway in 5 min., and a third because of a foot injury.

Apparatus.—The runway (Fig. 1) was $3\frac{1}{4}$ in. wide by 5 in. high and 18 ft. long; the last 2 ft. were not used. It was housed in a light-tight, sound deadened hull. The *S*s were run in the dark. The floor consisted of nine 2-ft. grid sections of $1\frac{1}{2}$ in. aluminum angle separated end to end by $\frac{3}{16}$ in., and side to side by $\frac{1}{4}$ in. Section 1 was divided into a $1\frac{1}{2}$-ft. startbox and a 6-in. starttime segment. The next five sections constituted the runway proper. Section 7 contained a 2-ft. delay chamber. This was offset 6 in. so that the last 6-in. part of Section 6 was included in the delay chamber and the last 6-in. part of Section 7 was used to measure starttime out of the delay-chamber segment. The next $2\frac{1}{2}$ ft. contained a 24-in. endbox. Eight times were measured, by means of electronic switches and .01-sec. Standard Electric timers: the time to traverse the first 6 in. out of the startbox, the times to traverse each of the next five 2-ft. sections, the time to traverse the first 6 in. out of the delay chamber, and finally, the time to traverse

---

[1] This investigation was supported by Research Grants M-817 and M-3328 from the National Institute of Mental Health, of the National Institutes of Health, Bethesda, Maryland. A report of these data was made at the 1959 Midwestern Psychological Association meetings in Chicago.

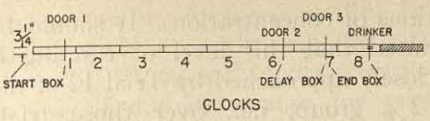


FIG. 1. Schematic of runway.

the endbox and contact the drinking tube. Time spent in retracing was included in the time to traverse the last section reached before the retracing began.

The startbox and the two delaybox doors were operated automatically. The startbox door opened 3 sec. following the closing of the startbox lid. The front delaybox door was down when *S* ran into the delay chamber and opened after a predetermined interval measured from the time *S* touched the seventh grid. The rear delaybox door came down when *S* touched the seventh grid, that is, 6 in. after he entered the delay chamber. Guillotine type doors operated by a motor driven screw were used.

A 50-ml. burette with a 4-mm. opening was used to deliver the reinforcement. It was located behind a 1-in. circular opening 2¼ in. above the floor in a partition in the endbox. The first contact with the tube started a Standard Electric timer. All subsequent licks were counted.

Pretraining.—One week before magazine training *S*s were put on a 22-hr. food and water deprivation schedule and were randomly assigned to the two concentration groups of 4% and 32% sucrose (percentage by weight).

The *S*s were adapted to *E* and the apparatus for 3 days just prior to the regular feeding time. Each *S* was handled 4 min. per day, then given a 5-min. exploration period in the runway. During exploration all doors were open.

Magazine training.—Each concentration group was randomly divided into two subgroups of 6 *S*s each. One subgroup of each concentration was trained and tested on alternate days. All *S*s were trained for 8 days to drink from a burette containing the appropriate solution.

Runway training.—Each *S* was given 2 trials per day in the runway on 21 alternate days. The delay door was down, forcing *S* to slow down or stop, and opened 1 sec. after *S* touched Section 7. The time allowed to drink was held constant at 20 sec. for 12 consecutive trials, 40 sec. for the next 8 trials, and 1 min. for the remaining 11 days of training and the remainder of the experiment. The time from the removal of *S* from the goalbox until he was placed again in the startbox, for his second daily run, was approximately 40 sec.

Delay.—The 4% and 32% *S*s were each matched into three delay groups of 4 *S*s each on the basis of their total running speed during the last six pretest trials. Three delays were used: 1 sec. (same as training), 3 sec., and 45 sec.

Two test trials per day were given each *S* on 3 alternate days. Half of the *S*s in each of the six groups were tested on different days; however, all *S*s continued running on the same alternate days as in training.

Retraining.—One week following the delay test trials *S*s were retrained to asymptotic runway speeds on 8 alternate days. Retraining conditions were identical to the pretest runway conditions except that the 1-sec. delay was eliminated.

Shift.—Two test trials were given each *S* on 10 alternate days. The *S*s for each concentration were matched into two groups on the basis of their total running speed during the last six retraining trials. The first group of each concentration was shifted to an 11.3% sucrose solution, whereas the second group was shifted to the opposite concentration (i.e., 4 to 32% and 32 to 4%).

RESULTS

Training

Total speeds of running (over the 10½ ft. from startbox door to delaybox) for the means of the 21 pairs of daily trials for the 4% and 32% concentration groups are presented in Fig. 2. An analysis of variance of these data (Table 1) shows that both the terminal speed and the rate of approach (Concentration × Trials interaction) to these speeds are func-

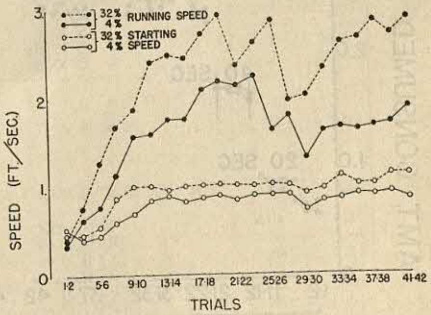


FIG. 2. Running speed (Clocks 1–6) and starting speed out of the delaybox (Clock 7) averaged over daily pairs of trials for the first 42 trials.

TABLE 1

ANALYSES OF VARIANCE OF TOTAL RUNNING SPEED (CLOCKS 1–6) AND TUBE CONTACT SPEED (CLOCKS 7–8)

Source	df	Running Speed		Tube Contact Speed	
		MS	*F*	*MS*	*F*
Between *S*s	23				
Concentration (C)	1	100.47	6.60*	7.63	7.43*
Error (b)	22	15.23		1.03	
Within *S*s	984	.70		.07	
Trials (T)	41	7.86	21.81**	.79	21.05**
C × T	41	.92	2.55**	.06	1.68*
Error (w)	902	.36		.04	

* $P < .05$.
** $P < .01$.

tions of concentration. It should be noted that this level of running is closely approached by Trial 12 of the 32% group, i.e., over those trials in which only 20 sec. contact time on the drinker was allowed.

In Fig. 2 is also presented the speed determined from the opening of the delaybox door (1-sec. delay) to the first contact with the drinking tube. An analysis (Table 1) of these data yielded results similar to those for total speeds.

In the first portions of Fig. 3 are presented the number of licks and the amount consumed for the two concentration groups. Analyses of these

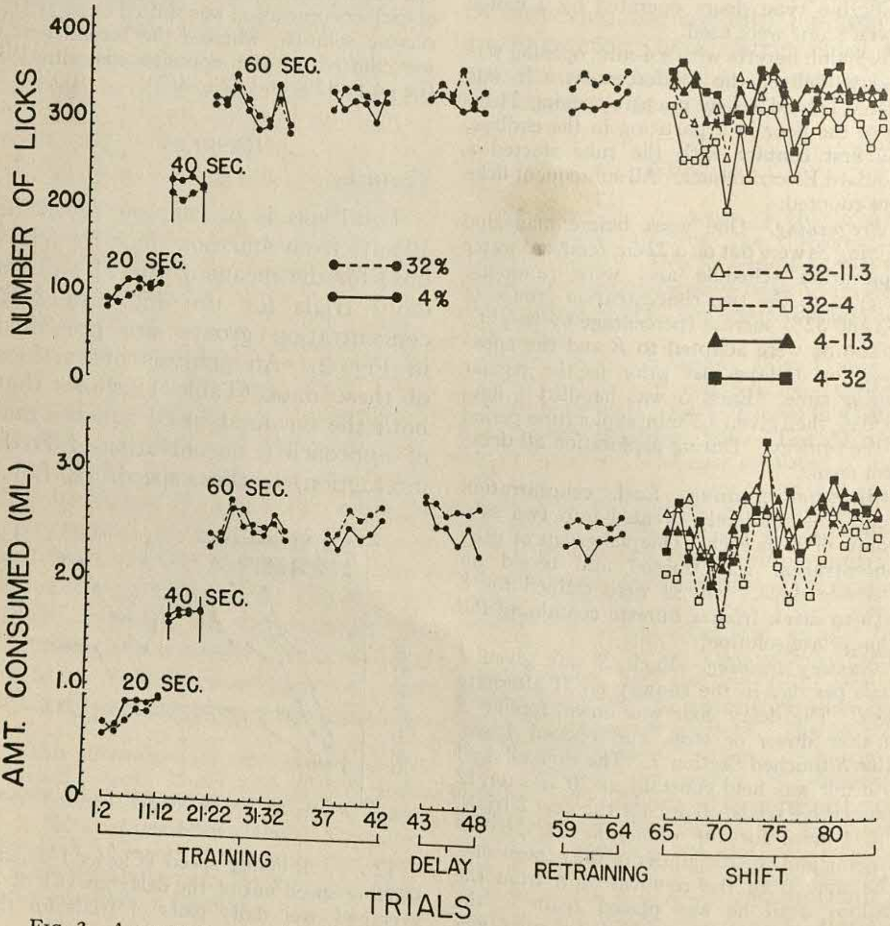


FIG. 3. Amount consumed and number of licks over the course of the experiment.

consummatory measures for the last eight trials at each drinking time (i.e., 20, 40, 60 sec. blocks) revealed no significant differences for trials within blocks, concentration, or their combination. Although in the early trials the 4% *S*s characteristically showed a greater mean number of licks. The differences between the numbers of licks and between the volumes consumed at the different drinking times (20, 40, and 60 sec.) were significant. Thus large differences in the rate of acquisition and terminal level of starting and running speed occurred with no discernible differences in total amount of consummatory activity. It can be seen from Fig. 3 that *S*s consumed approximately 2.5 ml. in 325 licks in 60 sec. of contact with the magazine. Any discussion of relative lick rates or ingestion rates of sugars of different concentrations must take the differences of viscosity into account. At 20° C., 32% sucrose is a little better than three times as viscous as 4% sucrose (Swindells, Snyder, Hardy, & Golden, 1958). This variable might account for the failure of the 32% groups to show a greater number of licks in the early trials. The data presented later in the present report show a progressive increase in the number of licks and volume consumed in 60 sec. so that the 32% groups eventually significantly exceeded the 4% on these two measures. Because of the differences in viscosity, lick rate and ingestion rate are not equivalent over a range of concentrations.

Delay

In order to assess the effect of delay the following analyses were performed on the six trials preceding the introduction of delay and the six delay trials.

Table 2 presents the mean starting speeds out of the startbox and the delay chamber (Clocks 1 and 7) for the three delay groups. It is clear from the analysis of variance of these means in Table 3 that the major effect of the introduction of the delay was on the delaybox starting times and produced a reduction in starting speed which was a function of the length of delay. The failure of delay to affect startbox starting speeds is to be expected since it usually has

TABLE 2

STARTING SPEEDS (FT/SEC) OUT OF THE STARTBOX (CLOCK 1) AND THE DELAYBOX (CLOCK 7) AS A FUNCTION OF DELAY

	Startbox			Delaybox		
	1 sec.	3 sec.	45 sec.	1 sec.	3 sec.	45 sec.
Predelay	.36	.34	.36	.41	.45	.42
Delay	.35	.35	.35	.39	.41	.24

TABLE 3

ANALYSIS OF VARIANCE OF SPEEDS FOR CLOCKS 1 AND 7

Source	*df*	*MS*	*F*
Between *S*s	23		
Concentration (C)	1	117.46	12.07**
Delays (D)	2	9.85	.32
C × D	2	20.90	.67
Error (b)	18	31.16	
Within *S*s	552	853.79	
Pretest-Test (E)	1	20.66	11.50**
Trials (T)	5	.60	.53
Clocks (V)	1	14.20	2.67
E × T	5	.42	.40
E × V	1	5.84	9.09**
E × C	1	23.35	2.02
E × D	2	10.42	5.80*
⋮			
D × E × V	2	7.69	11.99**
⋮			
C × D × E × T × V	10	1.26	1.02
Error (w)	414	1.28	
e_1 (CDES)	18	1.80	
e_2 (CDTS)	90	1.14	
e_3 (CDVS)	18	5.31	
e_4 (CDETS)	90	1.02	
e_5 (CDEVS)	18	.64	
e_6 (CDTVS)	90	.95	
e_7 (CDETVS)	90	1.23	

* $P < .05$.
** $P < .01$.

TABLE 4

Analysis of Variance of Speeds for Clocks 2–6

Source	*df*	*MS*	*F*
Between *S*s	23		
Concentration (C)	1	786.50	12.07**
Delays (D)	2	13.74	.21
C × D	2	6.39	.10
Error (b)	18	65.17	
Within *S*s	1416	2.52	
Pretest-Test (E)	1	15.72	2.50
Trials (T)	5	7.51	2.56*
Clocks (V)	4	71.25	31.24**
⋮			
V × C	4	17.23	7.56**
⋮			
C × E × T	5	7.87	3.01*
⋮			
C × D × E × T × V	40	1.95	1.01
Error (w)	1062	2.24	
e_1 (CDES)	18	6.28	
e_2 (CDTS)	90	2.93	
e_3 (CDVS)	72	2.28	
e_4 (CDETS)	90	2.61	
e_5 (CDEVS)	72	2.23	
e_6 (CDTVS)	360	2.08	
e_7 (CDETVS)	360	1.94	

* $P < .05$.
** $P < .01$.

its effect over trials. Concentration did not interact with any of the other variables.

Considering now the running speed between the startbox and the delaybox (Clocks 2, 3, 4, 5, and 6), neither the main effect of delay nor any of its interactions was significant (Table 4). Figure 4 shows the gradients of running, averaged over the two (pre- and postdelay) blocks of six trials.

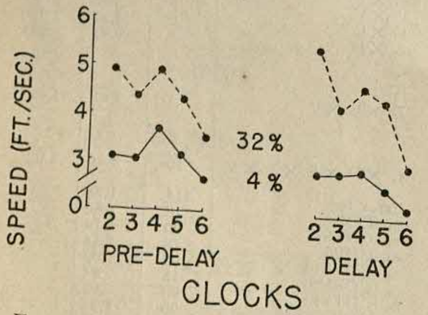


Fig. 4. Gradient of running to the delaybox before and after the introduction of delay.

TABLE 5

Speed of Approach (Ft/Sec) to Drinker as a Function of Delay and Concentration

Cond.	Trial	Delay					
		1 sec.		3 sec.		45 sec.	
		4%	32%	4%	32%	4%	32%
Predelay	6	1.14	1.95	1.50	2.98	1.62	2.95
Delay	1	1.46	2.33	1.89	2.45	1.00	1.09
	6	1.40	2.26	2.01	1.79	1.05	1.69

The fastest running occurred early in the runway. The gradient was steepest for the highest concentration. Again as with starting speed the failure of delay to influence the speed of approach to the delaybox is to be expected because of the small number of delay trials. It is interesting to note in Fig. 4 that the introduction of delay did not produce any immediate effect on the shape of the gradient of approach.

The major effect of the introduction of delay on the speed of approaching the drinker (Clock 8), as shown in Table 5, was to produce an *immediate* (first delay trial) drop in the speed of leaving the delaybox. An analysis of variance of these data is shown in Table 6. These results parallel those on the starting speeds from the delaybox (see Table 2). The size of this drop was an increasing function of the length of delay. Further, the greatest drop occurred for the 32% group, which showed the significantly greater speed, with little or no drop occurring for the 4% group. These data parallel those found by Collier and Marx (1957) for a delay introduced in the center of this runway.

Analyses of variance, (Table 7) of the volume consumed and of number of licks showed the interactions of pre- and postdelay blocks and trials and of trials and concentrations sig-

TABLE 6

Analysis of Variance of Speeds for Clock 8

Source	*df*	*MS*	*F*
Between *S*s	23		
Concentration (C)	1	25.97	11.90**
Delays (D)	2	7.43	3.40
C × D	2	.05	.02
Error (b)	18	2.18	
Within *S*s	264	.35	
Pretest-Test (E)	1	2.56	4.33
Trials (T)	5	.66	2.95*
.			
.			
.			
E × D	2	7.44	12.57***
.			
.			
.			
C × E × T	5	.59	2.67*
D × E × T	10	.28	1.28
C × D × E × T	10	.42	1.91
Error (w)	198	.26	
e₁ (CDES)	18	.59	
e₂ (CDTS)	90	.22	
e₃ (CDETS)	90	.22	

* $P < .05$.
** $P < .01$.
*** $P < .001$.

nificant for volume consumed (see Fig. 3). Both of these interactions appear to have resulted from a tendency, increasing over trials, for the 32% *S*s to consume more solution within the 1-min. drinking period. Delay appears to have affected the 4% groups, in that they showed a tendency to decline in consummatory activity following the introduction of delay (see Fig. 3). This pattern was reproduced in the lick data (Fig. 3), but did not reach significance.

Retraining

The data of the last 3 of the 8 retraining days (i.e., six trials) were analyzed. The starting speeds (Clock 1) differed only with respect to concentration. These times were, for 4% and 32%, respectively, .26 and .47 ft/sec. The analysis of running speed for Clocks 2–7 (with delay chamber now removed) showed Clocks ($P \leq .01$; $F_V = 32.26$; $df = 5/110$), Concentration ($P \leq .01$; $F_C = 71.84$; $df = 1/22$), and their interaction ($P \leq .01$; $F_{VC} = 27.51$; $df = 5/110$) to be significant. The differences between Trial 1 and Trial 2 on any given day, between days, and their interactions proved to be insignificant. A plot of the Concentration

TABLE 7

Analysis of Variance of Volume Consumed and Number of Licks

Source	*df*	Volume Consumed		No. of Licks	
		MS	*F*	*MS*	*F*
Between *S*s	23				
Concentration (C)	1	1.68	1.00	10,841.3	.79
Delays (D)	2	.43	.26	1,838.9	.13
C × D	2	1.17	.70	22,875.1	1.66
Error (b)	18	1.66		13,755.2	
Within *S*s	264	.09		1,306.6	
Pretest-Test (E)	1	.82	6.47*	4,087.6	1.38
Trials (T)	5	.07	1.08	2,057.5	1.89
E × T	5	.40	4.53**	1,138.5	.92
.					
.					
.					
T × C	5	.17	2.67*	892.4	.82
.					
.					
.					
C × D × E × T	10	.05	.51	1,019.6	.83
Error (w)	198	.08		1,325.1	
e₁ (CDES)	18	.13		2,960.6	
e₂ (CDTS)	90	.06		1,089.7	
e₃ (CDETS)	90	.09		1,233.2	

* $P < .05$.
** $P < .01$.

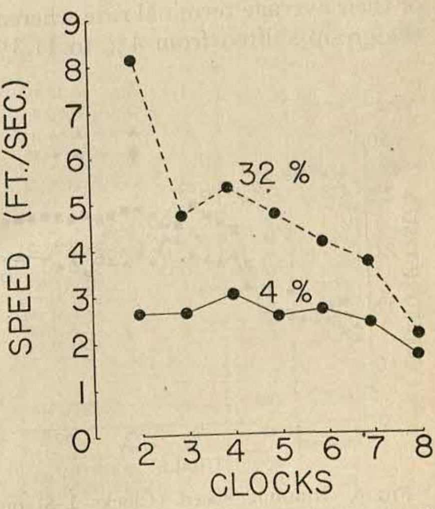


Fig. 5. The gradient of running to the endbox averaged over the last six retraining trials.

× Clocks interaction (including Clock 8) in Fig. 5 shows a steep gradient of running for the 32% *S*s, the fastest running occurring in the initial portion of the runway; and little or no gradient for the 4% group. This curve is similar to those shown in Fig. 4. The additional trials appear to have increased overall speed and steepened the gradient. An analysis of the drinker-approach speeds (Clock 8) showed concentration to be significant ($P \leq .01$; $F_C = 8.52$; $df = 1/22$). The analysis of volume consumed showed no differences (Fig. 3), while the analysis of number of licks showed a significantly larger number on the second trial of each day ($P \leq .01$; $F = 10.98$; $df = 1/10$), and a significantly greater number of licks for the 32% *S*s (Fig. 3) ($P \leq .01$; $F = 11.97$; $df = 1/22$).

Shift

Figure 6 details the effect over trials of the shifts in concentration of reinforcement on *total* running speed. Certain relations are readily apparent. By the third trial following shift, *S*s changed from 4% to 32% sucrose reached approximately three-fourths of their average terminal rate whereas the group shifted from 4% to 11.3% reached about four-fifths of their final rate. On the other hand, *S*s shifted from 32% to 4% or 11.3% showed a gradual decline in running speed and an interesting periodicity. The second trial of each day was performed at a lower speed than the first trial and the first trial of each day was higher than the last trial of the preceding day, with very few exceptions. This periodicity was most pronounced for *S*s undergoing the maximum downward shift (32% to 4%). The terminal rate for both groups of *S*s shifted to 11.3% was the same, approximately halfway between the preshift rates for the speeds on 4% and 32%. The *S*s shifted from 4% to 32% reached an asymptotic terminal speed above the 11.3% animals and below the preshift 32% speed. Periodicity, of the sort found for the shift-down groups, did not occur significantly in the total speed measures. The *S*s shifted from 32% to 4% arrived at a level below the 11.3% groups but above the preshift 4% animals. It is not clear whether this last group was asymptotic.

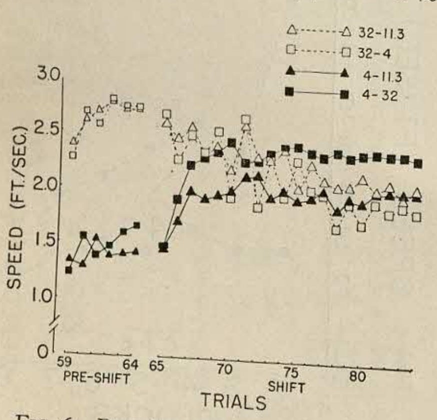


FIG. 6. Running speed (Clocks 1–8) on the six trials preceding and the 20 trials following shift in reinforcement.

In order to gain a more detailed look at these data, including the gradient of running, a series of analyses of various combinations of the shift groups were performed. The immediately preceding six *preshift* trials, the first six, and last six *shift* trials, the segments of the runway, and the first and second trials of each day were considered.

Comparisons of the 4% and 32% groups shifted to 11.3%.—Figure 7 presents the data for the groups shifted to 11.3%. Table 8 presents an analysis of these data for Clocks 2–7.[2] The first thing to be noted is

---

[2] The same variables as in the preceding analysis with the exception of the concentration groups were examined in the subsequent analyses. In order to save space

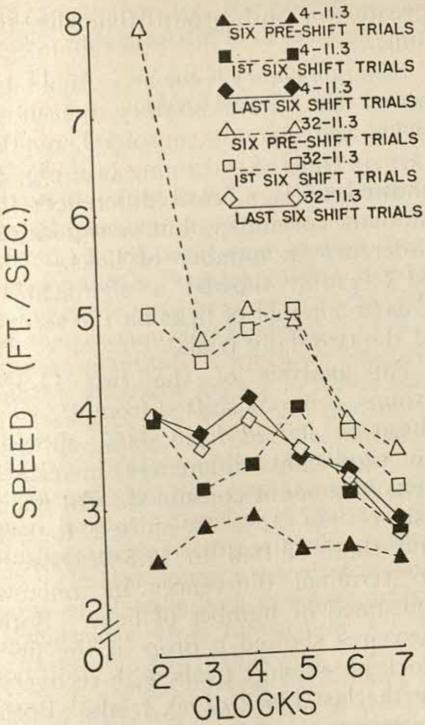


FIG. 7. The gradients of running on the six trials preceding, the six trials immediately following, and the final six trials following the shift in reinforcement.

that the steepness of the running gradient over the length of the runway was an increasing function, with the 11.3% gradient falling between the flat 4% gradient and the steep 32% gradient. As was the case with the shift in total speed, the 4%–11.3% group showed an almost immediate shift in gradient, while the 32%–11.3% group showed a slow shift to the 11.3% gradient. Most of the shift occurred in the last block of six trials. When Clock 1 and Clocks 2–7 were examined separately for the two shift-down groups (32%–11.3% and 32%–4%), it was found that the major part of the trial-to-trial periodicity shown in the total running speed measures was the result of starting speed measures rather than running speed measures. Rats started faster on the first trial than the second, but once running, ran at approximately the same speed on both of the two daily trials. The analysis on the speed over the last runway segment (Clock 8) indicates that the shift-up group's approach to the drinker was little affected by the shift, while the shift-down group showed a gradual decline to a slower approach to the drinker.

TABLE 8

ANALYSIS OF VARIANCE OF RUNNING SPEED OF THE 4%–11.3% AND 32%–11.3% SHIFT GROUPS

Source	*df*	*MS*	*F*
Between *S*s	11	72.76	
Concentration (C)	1	400.37	10.00*
Error (b)	10	40.00	
Within *S*s	1284	2.64	
Days (D)	2	.53	.34
Trials/Day (T)	1	2.78	1.27
Clocks (V)	5	76.07	20.64**
Pretest, Posttest$_1$, Posttest$_2$ (E)	2	15.36	2.61
⋮			
C × E	2	162.81	27.64**
C × V	5	20.92	5.66**
⋮			
C × V × E	10	13.18	3.86**
⋮			
C × D × V × E	20	1.93	2.05**
⋮			
C × D × T × V × E	20	1.22	.59
Error (w)	1070	1.93	
e_1 (CDS)	20	1.56	
e_2 (CTS)	10	2.18	
e_3 (CVS)	50	3.69	
e_4 (CES)	20	5.89	
⋮			
e_{10} (CVES)	100	3.41	
⋮			
e_{14} (CDVES)	200	.94	
e_{15} (CDTVES)	200	2.06	

* $P < .05$.
** $P < .01$.

---

neither the tables nor the significant effects are mentioned. No nonsignificant effect is discussed without this fact being noted. All significant effects are mentioned. Summaries of the designs may be had on request.

Comparisons of the preshift and postshift 4% and 32% groups.—Two analyses, one comparing the preshift 4% group with the postshift 4% (i.e., 32% to 4%) groups, and the second comparing the preshift 32% group with the postshift 32% groups (4%–32%), yielded some interesting comparisons. In both cases, the starting speed (Clock 1) was significantly different. The *S*s shifted from 32% to 4% never reached the slower starting speeds of the original 4% *S*s. Similarly, the *S*s shifted from 4% to 32% never obtained the speed of the original 32% *S*s. The shifted *S*s again showed the significant periodicity only on Clock 1. Considering the gradient of running (Clocks 2–7), neither of the 4% groups, the original or the postshift, showed a significant gradient, while both of the 32% groups did. Here again the *S*s shifted to 32% never obtained the level of speed of the original group nor did they show the steepness of gradient of running.

Comparison of the 32% groups shifted to 4% and 11.3% and of the 4% groups shifted to 11.3% and 32%.—An analysis of the two groups shifted from 32% (to 11.3% and 4%) showed that they differed significantly in the final gradients (Clocks 2–7) and in the rate of approach to their final gradient but that they did not differ significantly on Clock 8, although they both were significantly slower than their original speed. Periodicity was again significant only for starting speed. An analysis of the two groups shifted from 4% (to 11.3% and 32%) showed that they differed significantly in their final level, that on starting speeds those *S*s shifted to 32% ran faster on the second trial of each day's run while those shifted to 11.3% did not show this periodicity, and that the gradient of running was greater for both groups following the shift.

Consummatory behavior.—An analysis of the consummatory measures, licks, and amount consumed, on the last six trials of retraining (see Fig. 3) showed no significant differences for amount consumed, but a significant difference in number of licks. The 32% group showed a significantly greater number of licks on the second of the two daily trials.

An analysis of the two 11.3% groups, those *S*s shifted from 4% and those *S*s shifted from 32%, showed no significant differences in either licks or amount consumed. An analysis of the 32% *S*s shifted to 4% and those shifted to 11.3% showed no terminal differences in amount consumed or number of licks. Both measures showed a drop in the first block of six shift trials with recovery by the last block of six trials. Both measures also showed a significant tendency for less solution to be drunk with fewer licks in the first shift trials. An analysis of the 4% *S*s shifted to 11.3% and 32% showed that both groups consumed and licked more following the shift with greatest amount occurring by the last block of trials. Both groups showed a significant tendency to drink more and lick more on the second of the daily trials, with *S*s shifted to 32% drinking significantly more than those shifted to 11.3%.

Analyses comparing the preshift and shifted 32% groups showed only a significant increase in the number of licks on the second of the two daily trials. A similar comparison of the preshift and shifted 4% groups showed a significant tendency for the 4% *S*s shifted from 32% to make fewer licks and get more solution than the original 4% group. Further, the postshift 4% group consumed significantly more on the second of the daily trials than did their predecessors.

Discussion

If the consummatory response and its ancillaries are to be considered "responsible" for variations in performance of an instrumental response, as some accounts of reinforcement suggest (e.g., Estes, 1954; Premack, 1959; Spence, 1956), there must be some characteristics of the consummatory response, qualitative or quantitative, which are functions of the variables known to manipulate performance. In the present experiment two properties of the consummatory response, total frequency of licks and total amount consumed per trial, were examined as a function of the intensity (concentration) of the reinforcing stimulus. It was found that after a large number of trials both measures differed significantly as a function of the intensity of the reinforcing stimulus. Over a 1-min. interval of drinking *S*s licked faster and drank more of the 32% sucrose when compared with 4%. However, it was found that differences in starting speed and running speed evinced themselves very early, long before these differences in lick rate or amount consumed appeared,[3] and in fact the early differences that did appear were in the "wrong" direction. If then the technique of measurement of these variables can be considered to produce reliable results, it can be concluded that neither variations in the amount consumed nor the number of licks in 1 min. of reinforcement are *necessary* conditions for the production of differences in the instrumental response. This conclusion is reinforced by studies in which these variables are directly manipulated by varying tube size and amount consumed jointly (Knarr & Collier, 1960) or number of licks (Goodrich, 1960). It may be argued though that it is only differences in initial consummatory rate that are important. The results of the early trials in training require that this initial rate be sampled for less than 20 sec. It may not be concluded from these results that there is no property of the consummatory response which is necessary for differences in performance, but they seriously limit the possibilities.

The model of runway behavior which views it as a series of discrete responses pinned together by stimuli arising from the runway, from the deprivation operation, and from fragments of the consummatory response occurring to the preceding stimuli, leads naturally to the expectation of a gradient of running. A response chain is constructed from the terminal end, the responses most remote from the reinforcing stimuli being least strengthened. The present results, obtained under conditions where extraneous stimuli were minimized, show a gradient contrary to the "goal gradient" expectations. While no data were obtained for the very early trials, the picture obtained is of the evolution of a gradient with the fastest running occurring earliest in the runway. The steepness of this gradient is a function of variables which manipulate the effectiveness of the reinforcement, for example, shift in reinforcement magnitude. The failure to obtain a gradient accelerated in the direction of the endbox, under conditions where external stimuli are minimal, raises questions about the serial model previously employed to account for runway behavior. Functionally there appear to be only three units—starting, running, and stopping—which can be independently manipulated. This is shown by the effects of delays, shifts, etc., upon these measures. Starting appears to be a function of local stimuli, both extraneous and constant, and of the consequences of the chain. Starting appears to be, from these and prior data (Collier & Marx, 1957; Wist & Collier, 1960), more sensitive than running to amount, changes, and delay of reinforcement. An alternative to the

---

[3] The early differences were in the direction of slower drinking for the higher concentrations, perhaps reflecting viscosity differences. The development of a superior rate of drinking of 32% sucrose was relatively slow, considering the 8 days of preliminary magazine training.

classical view of running would be that it is a repetitive cyclic activity which is not chained by external stimuli but which is controlled by the terminal discriminative and reinforcing stimuli. It can be broken into segments by bringing it, at various points, under discriminative control, as is shown in maze studies or by the introduction of delays or other obstacles. Studies by Collier and Marx (1957) and Wist and Collier (1960) on the gradient of running preceding and following a point of delay have shown that such an obstacle has its effect mainly on the behavior immediately preceding and following the point and on starting behavior. Once running starts, it occurs at its usual (predelay) rate. Stopping appears to be a simple matter of preservation and to be under the discriminative control of the stimuli preceding the block or end of the runway.

The introduction of a delay results in an abrupt decrement in the performance in the running behavior following the delay point and a gradual decline over trials in the speed of running up to the delay point (Collier & Marx, 1957). It is apparent from the present results that there is no accompanying change in performance of the consummatory response as a function of the kind or durations of delay used, and that, therefore, disturbances in the consummatory response or its precursors can not be used to account for the disturbances in running.

The terminal level of responding and the course of the change are the major data in shift studies. No evidence for the "overshooting" of the preshift asymptotes as a function of shift in reinforcement was obtained in the present study. Rather, the residual effects of the rat's previous history were inertial. Both 11.3% groups, one shifted from 32% and one from 4%, reached a final common asymptote, whereas the 4% group shifted up to 32% reached a stable level below the previous 32% *S*s and the 4% group shifted from 32% reached a level of responding at the end of 20 trials above that for the original 4% group.

These remarks must be qualified by noting that no control groups, i.e., groups continued on 4% and 32%, are available for comparison. It is possible that the speed of running associated with these values of reinforcement converge, in which case the performance might meet or exceed the control performance. Such a substantial drift in performance seems unlikely on the basis of our previous runway experience. One major difference between this and other studies which have found shift effects of the opposite sort (e.g., Collier & Marx, 1959; Crespi, 1944; Zeaman, 1949) is that these *S*s had run 64 trials before the shift. Spence (1956) has shown a similar effect for his shift-up group after 48 trials. Unpublished studies from our laboratory show that the same variable reduces shift effects in the Skinner box. A second difference between this and other shift studies is that these *S*s were run two trials a day on alternate days. The failure to demonstrate over- or undershooting may be the result of the second trial which offers *S*s two samples of the change in reinforcement "close" together in time, differing in this respect from the studies of Crespi (1944), Zeaman (1949), and Spence (1956). That this variable has some effect can be seen in the periodicity associated with the two trials.

A surprising finding in the present study is that the course of change differed radically for the shift-up and shift-down groups. The shift-up was accomplished rapidly, the shift-down gradually and with a periodicity (in starting times) suggesting that the rat starts each day at a level commensurate with the previous reinforcement.

The data on the consummatory response did not parallel the running and starting data. The rats shifted from 32% to 4% showed increased variability and a dramatic drop to a level below the prior 4% *S*s with recovery over trials to the 4% level. The *S*s shifted to 11.3% became much more variable but showed an average level of responding not much different from their previous level. Both groups of *S*s shifted to

32% showed a gradual rise to a level of consummatory activity above the prior 32%.

On the basis of these findings it may be concluded that variations in the total amount of consummatory activity are not necessary conditions for reinforcement. Neither do they necessarily parallel variations in the instrumental activity.

SUMMARY

Two groups of rats were run in a 16-ft. runway to a drinking tube. One group received 4% sucrose and the other 32%. Both groups were allowed a fixed time on the drinking tube and the number of licks, amount consumed, and the times to traverse successive segments of the runway were measured. Following training, a delay, just preceding the endbox, was introduced. Following retraining after elimination of the delay, a shift in the value of reinforcement was introduced.

Variation in amount of consummatory activity over a fixed interval of time was not consistently associated with variation in running. The high concentration group always ran faster but drank slower (less) at first and faster (more) at the end of training. The introduction of delay did not result in variations in consummatory activity commensurate with those produced in running. The shift in reinforcement magnitude appeared to be mainly disruptive of consummatory activity, whether the shift was up or down. On the other hand, the shift-up groups showed a rapid increase in running speed to a level below that expected from previous asymptotic performance, while the shift-down groups showed a gradual drop to a level above that expected. The gradient of running to the endbox was such that under all conditions the fastest running occurred in the early portion of the runway. The steepest gradient occurred with the greatest magnitudes of reinforcement. The steepness increased over trials.

It was concluded that variation in total consummatory activity is not a necessary condition for variation in effectiveness of food reinforcement.

REFERENCES

COLLIER, G., & MARX, M. H. Frustration and performance. *Amer. Psychologist,* 1957, **12,** 458. (Abstract)

COLLIER, G., & MARX, M. H. Changes in performance as a function of shifts in the magnitude of reinforcement. *J. exp. Psychol.,* 1959, **57,** 305–309.

COLLIER, G., & MYERS, L. The loci of reinforcement. *J. exp. Psychol.,* 1961, **61,** 57–66.

CRESPI, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.,* 1944, **51,** 341–357.

ESTES, W. K. Analysis of the relation between amount of reinforcement and rate of conditioning in a bar-pressing situation. *Amer. Psychologist,* 1954, **9,** 361. (Abstract)

GOODRICH, K. P. Running speed and drinking rate as functions of sucrose concentration and amount of consummatory activity. *J. comp. physiol. Psychol.,* 1960, **53,** 245–250.

KNARR, F. A., & COLLIER, G. Taste and consummatory activity in amount of reinforcement effects. Paper read at Midwestern Psychological Association, St. Louis, April 1960.

PREMACK, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychol. Rev.,* 1959, **66,** 219–233.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SWINDELLS, J. F., SNYDER, C. F., HARDY, R. C., & GOLDEN, P. E. *Viscosities of sucrose solutions at various temperatures: Tables of recalculated values.* (Supplement to National Bureau of Standards Circular 440) Washington, D. C.: United States Department of Commerce, 1958.

WIST, E. R., & COLLIER, G. Delay and amount of reinforcement in the runway. Paper read at Midwestern Psychological Association, St. Louis, April 1960.

ZEAMAN, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.,* 1949, **39,** 466–483.

(Received August 22, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 496–502

FREE VERSUS SERIAL RECALL

NANCY C. WAUGH[1]

Harvard University

It has recently been shown (Waugh, 1961) that the serial memorization of digits proceeds at a constant rate, trial by trial. On each successive trial with a long list, the experienced *S* learns a new group of items farther down the list. Since the size of this group is constant, and since he seldom forgets an item once he has learned it, the total number of items that he can recall increases linearly with the number of trials. The curve for serial learning can in this case be considered to represent the number of trials multiplied by *S*'s span of immediate memory, up to the point where he masters the list.

With free recall, on the other hand, there is evidence that the learning curve is not linear but negatively accelerated (Bruner, Miller, & Zimmerman, 1955; Bush & Mosteller, 1955; Miller & McGill, 1952). Why this discrepancy? One obvious difference between the free and the serial recall experiment is the order in which *S* sees the items that he is to learn. Thus, in Bruner, Miller, and Zimmerman's (1955) study, the order of presentation was changed before every trial, while in tests of serial learning it is, of course, kept constant. Does the difference between free and serial recall depend on the order in which *S* sees the items on successive trials? Or does it depend instead on the order in which he is instructed to recall them? These alternatives were tested in the present study by comparing the learning curves obtained under the following three conditions: serial recall (S), free recall with the items to be recalled scrambled before every trial (F_v), and free recall with the order of the items kept constant (F_c). The items, monosyllabic English words, were presented and recited for a constant number of trials. The same group of *S*s served under each of the conditions, in order that the groups be directly comparable.

METHOD

The *S*s were nine college students, five men and four women, who were paid for their services. The material that they were to learn consisted of 18 series of words, each made up of 48 randomly ordered monosyllables. Half of these words were drawn from lists compiled by Egan (1948); the other half were drawn at random from a list compiled by Moser, Dreher, and Oyer (1957). No item ever appeared in more than one series.

For each of the three experimental conditions, a total of six series were exposed, each one for six trials. There should thus have been a total of 54 lists recited, six by each of 9 *S*s, under every condition. These numbers were actually 53, 53, and 52 for Cond. S, F_c, and F_v, respectively, due to absences by 2 *S*s.

The words were flashed onto a screen one at a time by means of a slide projector and an associated timing device. Each was exposed for 1.5 sec. and was followed by a blank interval of .5 sec. The total time between the onset of one word and the next was therefore 2 sec. After the series of 48 words had been shown, there was a 3-min. interval during which *S*s were to write down the words that they had just seen.

All *S*s learned the same lists in the same order. In every session, they recited two lists, each under a different experimental condition. The 18 lists therefore required nine sessions. The conditions were presented in the following order: F_c, F_v; S, F_c; F_v, S; F_c, S; F_v, F_c; S, F_v; F_c, F_v; S, F_c; and F_v, S.

---

[1] This work was carried out while the author was at the Lincoln Laboratory of the Massachusetts Institute of Technology.

Thus each was paired with each of the others three times; each appeared first within a session three times; and each appeared three times within each half of the experiment. There was a 15-min. interval between lists within each session.

The *S*s were instructed, under Cond. F_c and F_v, to recall as many words as they could in any order that they chose. Under Cond. S, they were instructed to recall the words in their proper serial positions (guessing, if necessary) and to write down early words before late ones. They were not specifically enjoined to attend to each successive word as it was exposed.

Each *S* was given two 6-page answer booklets, with 48 numbered blanks on each page, along with a copy of the general instructions, at the beginning of every session. The *S*s were forbidden to refer back to a given answer sheet after the next trial had begun, and were not allowed to write words down except during the interval between trials. The paper on which they wrote was quite thick, so that writing on one page did not produce legible impressions on the page underneath.

RESULTS

Serial recall.—The data obtained under Cond. S (serial recall) are summarized in Fig. 1. Here the lower set of points represents the mean number of words, out of a list of 48, that were recalled in their correct serial positions on each trial. The upper points represent the mean total number of words recalled per trial, both in correct and in incorrect serial positions. The smooth functions in Fig. 1 are best-fit straight lines through the origin, calculated for each set of data by a method of least squares. These lines clearly provide a good description of the data.

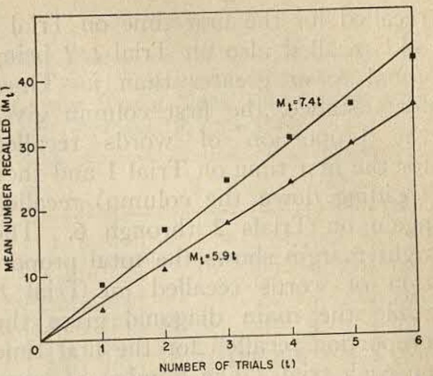


FIG. 1. Mean number of words recalled as a function of the number of trials, under Cond. S (serial recall). (The triangles denote the number recalled in correct serial positions, while the squares denote the total number recalled. The value shown for the total number recalled on Trial 6 is actually a median, rather than a mean: the distribution of scores on that trial was highly asymmetrical.)

Note that the number of words recalled in correct serial positions represents a constant proportion of the total number recalled. This total is thus the sum of two variables—the number in correct and the number in wrong serial positions—each of which is a linear function of the number of trials. It is reasonable to suppose that improperly placed words were actually remembered in the correct order relative to each other, rather than haphazardly, and were simply displaced on the answer sheet. No attempt has been made to test this hypothesis.

Table 1 shows the results of a detailed analysis of these data that was performed with the aid of an IBM 709 data processing machine. The input to this computer consisted of the number of each trial on which each *S* recalled each word in the proper serial position. Since there were six trials, there were 64 possible patterns of recall for a given word, ranging from recall on no trial to recall on all six. The data were reduced to a table showing the proportion of words associated with each of these 64 patterns, for all *S*s and all lists. These proportions were then combined to give the values shown in Table 1. Here each entry represents the proportion of words recalled for the first time on Trial *j* and recalled also on Trial *t*, *t* being equal to or greater than *j*. Thus, for instance, the first column gives the proportion of words recalled for the first time on Trial 1 and then (reading down the column) recalled again on Trials 2 through 6. The right margin shows the total proportion of words recalled on Trial *t*, while the main diagonal gives the proportion recalled for the first time on each trial. The number of cases on which these proportions are based (2544) is given by the number of subject lists (53) multiplied by the number of words per list (48).

The purpose of this analysis was to test the hypothesis that the curve of serial learning is linear because (*a*) words once learned are never forgotten and (*b*) a constant number of new words are learned on every trial, so that the total number of words recalled after a given trial is simply this constant number multiplied by the number of the trial. According to this hypothesis, therefore, the entries in Table 1 should all be identical, while the marginal totals should increase linearly. Obviously the hypothesis is somewhat too simple.

It appears, for one thing, that people sometimes forget a word after they have recalled it once. They are likeliest to do so on the very next trial, and are less likely to do so on later trials. This effect is evident in the columns of Table 1. Here we see, for example, that, on the average, about six words (12.4% of 48) are recalled for the first time on Trial 2. On the next trial, only about five of these six words are remembered. Then the proportion recalled increases gradually, until the original six words are again almost always recalled on Trial 6. This increase is consistent with the well-known principle that the more times a word has been recalled already, the likelier it is to be recalled again.

TABLE 1

Proportions of Words Recalled on Each Trial (t), Analyzed According to the Trial on Which They Were First Recalled (j): Cond. S ($N = 2544$)

t	j						Total
	1	2	3	4	5	6	
1	.105						.105
2	.096	.124					.221
3	.099	.107	.167				.373
4	.103	.116	.147	.143			.509
5	.101	.120	.147	.122	.127		.617
6	.103	.123	.157	.132	.117	.103	.735

It is also apparent that the number of words recalled for the first time on each trial is not perfectly constant. If we follow the main diagonal of Table 1, we see instead that this number, expressed as a proportion, increases up through Trial 3. It then decreases, so that by Trial 6 the proportion of words recalled for the first time is almost exactly what it was on the first. The departure of these six proportions from equality is highly reliable ($\chi^2 = 58$; $P < .001$). The direction of this departure, furthermore, is contrary to what we might have expected. Retaining words already learned does not seem to interfere with learning new ones. Rather, it appears that mere exposure to the individual items in a list may facilitate the forming of associations between them, an effect that has been reported by Hovland and Kurtz (1952) and by Noble (1955). Note that the two effects evident in Table 1—the forgetting of old items and the recalling of more new ones on intermediate trials—are complementary, so that the total proportion of items recalled does increase linearly with the number of trials.

Free recall.—Free recall presents a very different picture. Figure 2 shows the mean number of words, estimated across *S*s and lists, that were recalled under Cond. F_c (constant order of presentation) and F_v (variable order). These values, denoted by the circles and the triangles for Cond. F_c and F_v, respectively, have been fitted with a simple negative exponential curve. A striking feature of Fig. 2 is the extreme similarity of the two sets of data. None of the small differences between corresponding pairs of points are statistically reliable. The inference from these data is either that scrambling the words in a list before every trial does not retard learning, or that keeping them in the same order does not facilitate it, when learning is tested by free recall. Repeated contiguity between the items to be learned is in this instance an unimportant variable.

Let us now compare Fig. 2 with Fig. 1. It is evident that, on early trials, more words are recalled, on the average, under Cond. F_c and F_v than under Cond. S—a fact consistent with Deese's (1957) data on the oral recall of spoken words after one trial. Note, however, that by Trial 6 the total number recalled is greater when recall is serial than when it is free. An appreciable number of the lists were, as a matter of fact, recited perfectly on this trial under Cond. S, whereas almost none of the lists were ever mastered under Cond. F_c and F_v. Serial recitation, therefore, was eventually the more effective method of study in this experiment.

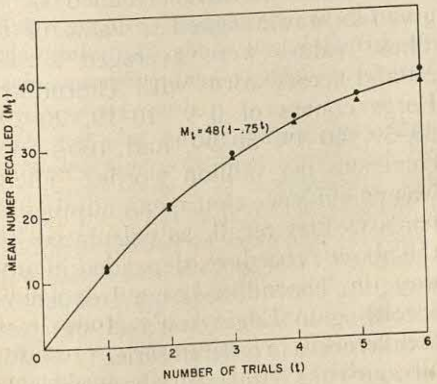


FIG. 2. Mean number of words recalled as a function of the number of trials, under Cond. F_c and F_v (free recall). (The circles represent Cond. F_c—words presented in a constant order; the triangles represent Cond. F_v—words presented in a variable order.)

TABLE 2

PROPORTIONS OF WORDS RECALLED ON EACH TRIAL (t), ANALYZED ACCORDING TO THE TRIAL ON WHICH THEY WERE FIRST RECALLED (j): COND. F_c ($N=2544$)

t	j						Total
	1	2	3	4	5	6	
1	.237						.237
2	.208	.244					.452
3	.205	.207	.185				.597
4	.214	.204	.151	.136			.705
5	.208	.207	.154	.112	.089		.770
6	.221	.219	.163	.116	.074	.045	.838

The data on free recall are further analyzed in Tables 2 and 3, which are analogous to Table 1 and were constructed in the same way. Here again each entry represents the proportion of items recalled for the first time on Trial j and also recalled on Trial t, for $t \geq j$. A comparison of Table 2 with Table 3 reveals differences of very small magnitude. Neither the rate at which old items are forgotten nor the rate at which new ones are acquired appears to depend in any way on whether the list is presented in a constant or in a different order on every trial. Tables 2 and 3 show also that the proportion of new words recalled on each trial decreases with the number of the trial, from an average of .249 on Trial 1 to an average of .048 on Trial 6.

If we now compare Tables 2 and 3 with Table 1, we see that relatively more words seem to be forgotten when recall is free than when it is serial. (Thus, for example, about

TABLE 3

Proportions of Words Recalled on Each Trial (t), Analyzed According to the Trial on Which They Were First Recalled (j): Cond. F_v ($N=2496$)

t	j						Total
	1	2	3	4	5	6	
1	.261						.261
2	.215	.233					.448
3	.218	.181	.188				.587
4	.229	.190	.143	.126			.688
5	.233	.195	.159	.093	.077		.757
6	.229	.205	.162	.098	.057	.051	.802

10% of the words recalled on Trial 1 under Cond. S are forgotten on Trial 2, while under Cond. F_c and F_v an average of about 16% of the words recalled on Trial 1 are forgotten on the next trial.) This inference was tested by the following procedure. The number of words recalled for the first time on Trial j and either (a) recalled again on Trial $j+1$ or (b) forgotten on that trial, under (a) Cond. S or (b) Cond. F_c and F_v, were combined in a 2×2 contingency table. Five such tables were constructed, one for every j from 1 to 5. The value of chi square then calculated for each table was in every case statistically significant ($\chi^2 = 7.45, P < .01$; $\chi^2 = 4, P < .05$; $\chi^2 = 16.63, P < .001$; $\chi^2 = 7.01, P < .01$; and $\chi^2 = 24.49, P < .001$, for $j = 1$, 2, 3, 4, and 5, respectively).

Both of the effects evident in Tables 2 and 3—the forgetting of recently learned words, the acquiring of fewer new ones on later trials—influence the total number of words recalled on each successive trial. The overall rate at which the list is learned thus decreases with the number of trials. The result is a negatively accelerated learning curve, in contrast to the linear curve that results when recall is serial. The difference between these curves depends not on the order in which the items are presented but rather on the order in which the learner must recall them.

The effect of word frequency.—Since the various lists used in this experiment were made up of words selected at random, it was assumed that each would contain both common and unusual words in about the same proportions. The effects of word frequency on speed of learning should thereby have been spread uniformly over the three experimental conditions. One may nevertheless wonder about the extent to which *S*'s previous familiarity with a word did affect the likelihood of his recalling it in this experiment.

In order to answer this question, the 864 words used in this study were sorted according to their frequency of usage in printed English, as reported by Thorndike and Lorge (1944). A count was then made of the number of trials elapsing before each word was recalled for the first time by each *S*. Thus, a word first recalled on Trial 1 by 1 *S*, and on Trial 6 by another *S*, was assigned values of 0 and 5, respectively. A word that was never recalled by a given *S* was assigned a value of 6. These values were averaged across *S*s and across words with Thorndike-Lorge counts of 0–9, 10–19, 20–29, 30–39, 40–49, 50–99, and 100+ occurrences per million words. There was no evidence that mean number of trials to first recall, as calculated by the above procedure, depended in any way on Thorndike-Lorge frequency, according to Edgington's (1961) test of differences in ordered series ($P = .56$, for four runs and seven observations).

Several earlier investigators have studied ease of learning as a function of word frequency. They have variously reported a positive monotonic relation (Hall, 1954), a U-shaped function (Nachmias, 1958; Peters, 1936), and no relation (Underwood & Postman, 1960). The results of the present study agree with those of the latter. The procedures used in these various studies differ so widely, however, that disparate results are not necessarily contradictory.

Discussion

The data observed here on the serial learning of words are consistent with those reported previously on the serial learning of digits (Waugh, 1961): the number of items recalled in correct serial positions increases linearly with the number of trials. While the number of new words learned on each trial is not perfectly constant, but instead increases somewhat as *S* becomes familiar with the list, and while some forgetting does take place, the linear learning curve might nevertheless be said to represent a cumulation of memory spans. When recall is serial, the *S*'s usual strategy is to overlearn successive groups of words, starting at the beginning of the list, and to ignore later items while doing so. Thus 1 *S*, for instance, reported that he would "grab a mindful" of new words on every trial, rehearse this group, and then attach it onto the portion of the list that he already knew. The question that remains to be answered is just why this "mindful" that is assimilated on each successive trial should be limited to five or six words.

It may be an important fact that serial learning was here tested not by the usual method of anticipation but rather by what might be termed a modified method of complete presentation (McGeoch, 1942, p. 10). The *S* was not required to respond to each word as it came up but was instead free to distribute his attention as he pleased throughout the exposure trials. He could thus either ignore or overlearn whichever items he chose. It is possible, therefore, that the linear learning curve is an effect of the testing procedure used here. If a constant number of items were attended to and thus learned completely and discontinuously on each trial, while others remained unnoticed (and thus unlearned), the result would be a linear learning curve. In this connection, it is interesting to note that several recent studies which are said to demonstrate the all-or-none learning of paired associates have employed an exactly analogous testing procedure (Clark, Lansford, & Dallenbach, 1960; Estes, Hopkins, & Crothers, 1960; Rock, 1957). Now a method of anticipation would presumably require *S* to shift his attention to each successive item in a list, regardless of whether he had thoroughly committed the previous one to memory. Would either serial or paired-associate learning still appear to be discontinuous when tested by this method?

The major reason, of course, for not using a method of anticipation in the present investigation was to make the conditions of serial recall comparable to those of free recall. Obviously, it would not make sense to test free recall by a method of anticipation. The present data show that when all of the experimental conditions except the instructions to *S* are equivalent, free recall differs markedly from serial. In the former case, the learning curve is negatively exponential; in the latter, it is linear.

The present investigation has shown also that it does not matter whether a series is reordered before every trial or whether it is presented in a constant order, when recall is free. This surprising outcome implies that repeated contiguity between words cannot be an important principle of verbal learning when *S* is not required to associate contiguous words. Contiguity may perhaps be necessary or even sufficient for the forming of associations. It does not seem to be necessary, however, for the maintaining of associations. A word is recalled just as easily when it appears in a new context as when it reappears in the same context on every trial. Of course, it may be that multiple associations accrue to words that are scrambled before every exposure. If so, the present findings argue that these additional associations neither facilitate nor inhibit recall.

An alternative way to interpret these findings is to assume that no associations are formed between any of the items in a series when recall is free, so that contiguity is in this case a completely irrelevant variable. The persistent occurrence of clusters of words on successive recall trials under Cond. F_V, evident on visual inspection of the *S*s' answer sheets, renders this assumption an unlikely one.

SUMMARY

The present study compares the free and serial recall of lists of monosyllabic English words over the course of six trials. In agreement with earlier findings, the learning curve for serial recall is linear, while the curve for free recall is negatively accelerated. The difference between these functions results from the fact that fewer new words are learned on each successive trial when recall is free, while an approximately constant number are learned when recall is serial. Relatively more words are forgotten on each trial, furthermore, when recall is free than when it is serial. Ease of learning, as denoted by the number of the trial on which a particular word was first recalled, is independent of this word's Thorndike-Lorge frequency under both conditions of recall.

Free recall was not found to depend in any way on whether the words to be learned were scrambled before every trial or were presented in the same order on all trials. The difference between free and serial learning must therefore result from the order in which the *S* is required to recite these words, rather than on the order in which he sees them. Since presenting a list in the same order on every trial does not facilitate its free recall, repeated contiguity between items to be learned does not appear in this case to be an important variable. It is concluded that serial learning represents a cumulation of memory spans. Free recall, on the other hand, is not so easily interpreted.

REFERENCES

BRUNER, J. S., MILLER, G. A., & ZIMMERMAN, C. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, **49**, 187–192.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

CLARK, L. L., LANSFORD, T. G., & DALLENBACH, K. M. Repetition and associative learning. *Amer. J. Psychol.*, 1960, **73**, 22–40.

DEESE, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

EDGINGTON, E. S. Probability table for number of runs of signs of first differences in ordered series, *J. Amer. Statist. Ass.*, 1961, **56**, 156–159.

EGAN, J. P. Articulation testing methods. *Laryngoscope*, 1948, **58**, 955–991.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

HALL, J. F. Learning as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 138–140.

HOVLAND, C. I., & KURTZ, K. Experimental studies in rote-learning theory: X. Prelearning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.

MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.

MILLER, G. A., & MCGILL, W. J. A statistical description of verbal learning. *Psychometrika*, 1952, **17**, 369–396.

MOSER, H., DREHER, J. J., & OYER, H. J. *One-syllable words.* Columbus: Ohio State University Research Foundation, 1957.

NACHMIAS, J. The effect of stimulus-heterogeneity on free recall. *Amer. J. Psychol.*, 1958, **71**, 578–582.

NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

PETERS, H. N. The relationship between familiarity of words and their memory value. *Amer. J. Psychol.*, 1936, **48**, 572–584.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

WAUGH, N. C. Length of series and the learning curve. *Amer. J. Psychol.*, 1961, in press.

(Received August 22, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 503–509

THE EFFECTS OF AMOUNT OF REWARD, REQUIREMENT, AND SEVERAL RELATED PROBABILITIES ON HUMAN PERFORMANCE[1]

SHEILA G. ZIPF

University of California, Berkeley

The present study is an investigation of change in human performance as a result of variations in motivating conditions. Three of the four variables whose arrangement constituted the motivation conditions have been shown to be positively related to performance in animals. They are: amount of reward (e.g., Crespi, 1944; Hulse, 1958), probability of reward for a given performance (Hulse, 1958; Lewis & Cotton, 1957), and amount of work required in order to obtain the reward (Crutchfield, 1939). The fourth variable, probability of success, has been of great interest in recent human research but has been used to predict choice rather than performance.

A number of theorists have suggested that performance is a result of some combination, usually multiplicative, of amount or value of the reward and probability or expectancy of reward (Atkinson, 1957; Edwards, 1955; Lewin, 1938; Rotter, 1954; Tolman, 1955). Feather (1959a) reviewed these positions in a recent article. It is frequently overlooked, however, that the probability of reward (PrR) may be independent of the probability that the organism will be successful at the required task (PrS). In general, PrR is fixed by the situation while PrS is a result of the organism's ability at the task. Both of these probabilities are sufficiently meaningful for human *S*s to report what they think they are and for each of these reports to correlate significantly and independently with change in performance (Zipf, 1960). As to the combination of these two probabilities, it is reasonable to expect that they multiply together. This joint probability may be thought of as the probability of obtaining the reward if the organism attempts the task.

Although usually held constant, it is clear that within limits, the more work that is required to obtain a goal, the more is done. This dimension might well interact with amount of reward, PrR and PrS. The experiment was, therefore, designed to observe the effects and interactions of these four variables.

METHOD

Design.—A 3 × 2 × 2 × 2 factorial design was used. PrS was given three levels (5%, 30%, 95%) because of some indication from earlier data that a PrS of around 30% produced less change in performance than lower or higher probabilities. The rewards offered were either 10¢ or $1. The PrR was set at 50% or 90%. Subjects were required to go 1.1 or 2 times as fast as they had been going in order to obtain the reward. A positive relationship to change in performance was expected for each variable.

Because this design diverges considerably from most studies of performance, the differences require clarification. A primary difference is that the amounts, probabilities, and requirement for reward were not learned over time through experience but rather were given directly by instructions. This method is often used in human studies of choice.

---

[1] This investigation was supported by a postdoctoral fellowship (MF-9450) from the National Institute of Mental Health, United States Public Health Service and was conducted at the University of Rochester. The writer wishes to express appreciation to Vincent Nowlis, who sponsored and encouraged this research, and to the Department of Psychology, University of Rochester, for the facilities provided.

Speed on the task was the performance to be considered and the reward was offered for a particular speed. This differs from the usual animal study in which the reward is given for getting to the end of the runway or for making the correct choice, and the speed which is observed does not affect whether the animal obtains the reward or not. Therefore, the relationship of motivation to speed of performance is more direct in the present situation.

The effect of motivation on performance vs. learning was not of concern here, and *S*s were not run to an asymptote prior to changing the motivational conditions. Since opportunity to learn is the same in all conditions, it does not confound the results. It will lead to inaccuracy if different *S*s are more or less able to change their speed because they are at different stages of learning. The following procedures were designed to reduce effects of learning and of individual differences. A simple task was chosen. The *S*s were told how to perform the task and asked to do it that way only. Everyone had 12 min. of practice prior to the test trials. Since the details of amount of reward, probabilities, and requirement were given in instructions, they did not have to be learned. Furthermore, absolute scores on the task are not used. Rather, each *S*'s score is his speed following the change in motivating conditions divided by his speed prior to that. This index is intended to control for differences in ability and initial motivation.

Task.—A card sorting task provided the measure of performance. The *S*s were given a deck of IBM cards having from 16 to 25 holes each, and a rack with 10 slots labeled 16, 17, . . . 25. The deck of cards was in a box which had a marker on the back that could be used to indicate how many cards had been removed. The task was to count the number of holes in each card and place it in the appropriate slot in the rack. The *S*s were asked to take about 10 cards in hand at a time, place one card face down on the table, and count the holes by twos. They were told that it was important to the experiment that it be done this way. Each frequency of holes was punched in three different patterns so that the frequency could not be recognized from the pattern. The decks which *S*s received were thoroughly mixed for frequency and pattern. The task had been used in several earlier experiments (French, Morrison, & Levinger, 1960; Zipf, 1960).

Procedure.—Subjects were assigned to sessions in groups of size 2 to 12. The task was explained and they were told that they might earn some money later. They sat in separate booths. There were three work periods of 6 min. each, with breaks between each minute and a longer break between work periods. During these breaks, different colored cards were placed in each slot so that work could later be identified as to the minute in which it was done. The markers were set to zero before each 6-min. work period. With them, *S* was able to see approximately how many cards he had sorted during the work period.

Before *S*s started to work, they were told to sort at a comfortable speed and to be accurate. Before the second work period, they were given an instruction sheet telling them to be accurate and sort at a speed that was comfortable, but reasonably fast. After the second work period, *E* went around to each booth, looked at the marker behind the deck of cards, and gave each *S* an instruction sheet (see below). The *S*s were randomly assigned to the 24 conditions described by these instructions. A brief questionnaire followed the instruction sheet, and another was filled out following the third work period.

Instruction sheet.—The following is a copy of the instruction sheet used before the third work period. All of the alternatives are indicated, but only one appeared in each position on the actual sheets. The italicized portions were filled in by *E* while standing beside *S*.

> Last time you sorted about *X* cards. In the next work period, which will also be six minutes long, I would like you to try to sort one and one tenth (two) times as many as that correctly. That is, I would like you to try to sort *N* cards correctly next time.
>
> If you had about average accuracy in the last work period, from what I know about how others have done with this amount of practice and under these conditions, your chances of being able to sort the requested number of cards next time are *very low* (*low*) (*high*). That is, the probability that you will be able to count *N* cards correctly next time is about *5* (*30*) (*95*) %.
>
> If you do sort that many cards correctly next time, you may receive a reward of 10¢ ($1) before you leave. If you do sort the requested number of cards cor-

rectly, I will give you a folded slip of paper with a number from 1 to 10 on it. You may have five (nine) guesses at the number written on it. If you guess it, you will receive the dime (dollar). This means that you have a 50 (90)% chance of getting the dime (dollar) if you count the requested number of cards correctly. Only cards that are sorted correctly will be counted in figuring your total.

After you have answered the questions on the next page: keep this sheet where you can see it so you will remember what it says. Then, adjust the marker paper behind the cards so that the top line on the paper is even with the top of your stack of cards. This will show how many cards you are counting as you go along.

Subjects.—The *S*s were 289 men and women volunteers from freshman English and introductory psychology classes at the University of Rochester. Men and women were always in separate sessions. They were told that the purpose of the experiment was to study the effects of different practice and working conditions on routine tasks. Five *S*s had to be dropped when their performance data were lost through some error by *E* or *S*. To equalize cell *N*s, only 10 *S*s per cell could be used. The first six males and four females who had been obtained in each condition, 240 *S*s in all, were used in analysis of the data.

RESULTS

An analysis of variance was done on the ratio: total number of cards *S* sorted in the third work period divided by total number of cards *S* sorted in the second work period. The only significant effect was required speed (Req). Since experimental instructions were quite complicated, it seemed likely that the variables would have their effect most clearly immediately after the instructions were given. The ratio: total number of cards sorted in the first minute of the third work period divided by total number of cards sorted in the first minute of the second work period ($t_{3,1}/t_{2,1}$) was computed. The abbreviation $t_{i,j}$ means the total number of cards sorted in the jth minute of the ith work period. These scores were transformed to $\log(1+X)$ which corrected a significant heterogeneity of variance. The results of analysis of variance on these transformed scores are given in Table 1. Probability of reward and Req each have significant effects. Probability of success is also significant, but the effect is opposite to what was expected. Amount of reward is not significant, nor are any of the interactions. One second-order interaction which was just significant using the untransformed scores is just below significance with the transformed scores. This interaction appears to be due to the irregularity of effects in Low Reward as compared to High Reward conditions.

Since there are no interactions, the variables appear to be additive rather than multiplicative. At least one animal study (Hulse, 1958) failed to obtain an interaction of two of these variables: amount of reward and PrR. However, it remains possible that better controlled research with a wider range of the independent vari-

TABLE 1

ANALYSIS OF VARIANCE OF TRANSFORMED PERFORMANCE SCORES: LOG ($t_{3,1}/t_{2,1}+1$)

Source	*df*	*MS*	*F*
Amount of reward	1	000	—
Probability of reward (B)	1	.026	8.67**
Probability of success (C)	2	.013	4.33*
Requirement (D)	1	.019	6.33*
A × B	1	.006	2.00
A × C	2	.002	.67
A × D	1	.002	.67
B × C	2	000	—
B × D	1	.002	.67
C × D	2	.001	.33
A × B × C	2	000	—
A × B × D	1	000	—
A × C × D	2	.009	3.00
B × C × D	2	000	—
A × B × C × D	2	.004	1.33
Error	216	.003	

*$P < .05$.
**$P < .005$.

ables will support the multiplicative hypothesis.

Amount of reward.—As is often the case with human *S*s, need or desire for the incentive was not controlled in this study. It is thus possible for variability in desire for the reward to obscure the effect of amount of reward. Using verbal reports, we can infer this from these data. Immediately after being informed of the reward conditions, *S*s were asked how much they wanted the reward. Answers were checked on a nine-point scale. Table 2 shows the distribution of high and low reward groups on this dimension broken as nearly as possible into thirds. Although the larger reward is generally more preferred, there are many who want 10¢ more than others want $1. Presumably the dime is valued as a symbol of success. Desire for the reward has a low but significant correlation with the transformed performance scores across all conditions ($r = .17$, $N = 214$, $P < .05$).

TABLE 2

NUMBER OF *S*S IN HIGH AND LOW REWARD GROUPS WITH VARYING DESIRE TO OBTAIN THE REWARD

Amount of Reward	Desire for Reward			Total
	Low	Medium	High	
10¢	52	38	30	120
$1	17	33	70	120
Total	67	71	100	240

Another matter of interest concerning the amount of reward is the general regularity of the data in High and Low Reward. The other three variables are all closely connected to the monetary reward. PrR is clearly so. Req is the speed necessary to obtain the reward. PrS is the probability of reaching the Req in order to be eligible for the reward. Assuming that 10¢ is generally a less valuable reward than $1, even though it produces as much work by *S*, we might expect that relatively more nonmonetary factors are motivating *S* to work in Low Reward conditions than in High Reward. Need to achieve, desire to please the experimenter, and other personality factors may be involved. Regardless of what the other factors are, we would expect that PrR, Req, and PrS would be more important determiners of performance in High Reward than in Low. Table 3 presents the means of the transformed performance scores for the entire 24 cells. It is readily seen that within High Reward these three variables each have completely regular effects in line with the overall findings, while within Low Reward

TABLE 3

MEANS OF TRANSFORMED PERFORMANCE SCORES IN EACH CONDITION

Req	PrR	10¢ Reward			$1 Reward		
		5% PrS	30% PrS	95% PrS	5% PrS	30% PrS	95% PrS
2.0	50%	.320	.320	.326	.367	.334	.320
	90%	.337	.369	.345	.373	.337	.327
1.1	50%	.328	.313	.280	.318	.316	.302
	90%	.364	.326	.330	.340	.336	.311

Note.—N = 10 per cell.

these variables have quite irregular effects.

Probability of reward.—In comparing the experimental groups on initial performance ($t_{2,1}$), it was found that those *S*s who later received the Low PrR instructions were significantly faster than those who later received the High PrR instructions. All *S*s had been treated alike before these instructions were given. They were randomly assigned to conditions and in each experimental session every *S* was assigned a different condition. It appears to be a chance effect, but, it makes interpretation of the PrR variable questionable. However, it is possible to draw a tentative conclusion. On the absolute test score ($t_{3,1}$) High and Low PrR do not differ. They have shifted from a pre-experimental difference to no difference, indicating that the High PrR *S*s increased more than the Low PrR *S*s, as expected.

As shown on Table 1, a significant effect is obtained for this treatment using the index $t_{3,1}/t_{2,1}$, which is a measure of relative change. However, this index takes advantage of the initially low score of the High PrR *S*s. The index is negatively correlated to $t_{2,1}$($r = -.41$, $N = 240$, $P = .01$). This is precisely the type of correction which the index is designed to make, giving more weight, for the same absolute increase by initially low scorers, than to initially high scorers. A more appropriate index would, of course, be independent of initial scores. Neither the assumptions made by the index nor the cause of the initial difference in the experimental groups are sufficiently understood to accept this result. Current research, utilizing more levels of PrR, indicates that change in performance is an increasing linear function of PrR.

Probability of success.—This variable did not show the relationship to performance that was expected. A trend analysis is not presented because of the unequal spacing of the levels. However, performance change is clearly a decreasing function of PrS. The trend appears to deviate slightly from linear, having the general form of a negative exponential function. (The means of transformed scores for 5%, 30%, and 95% are respectively: .341, .331, .318, $\sigma_{\bar{x}} = .006$.) The direction of this relationship is opposite to that found in earlier correlation studies. The lower the probability of success, the greater was the increase in speed. A measure of what *S* thought his probability of success was showed that the probability given in the instructions was generally accepted. However, the instructions seem merely to have indicated to *S* how hard he would have to work to achieve the goal. Right after reading the instructions about the conditions of reward, *S*s were asked how hard they would have to try in order to succeed. Using this measure as the dependent variable, an analysis of variance of the same form as that shown in Table 1 was done. Both Req and PrS had significant effects (Req, $df = 1$, $F = 121.14$, $P < .005$; PrS, $df = 2$, $F = 6.18$, $P < .005$). No other effects were significant. The larger the requirement and the less the purported chances of success, the harder *S*s felt they would have to try to succeed, and, indeed, the more they increased their speed.

The dependent variable.—It was decided prior to obtaining the data for this experiment that the dependent variable would be a ratio of speed on the test trial divided by speed on the trial prior to the test. This index had been used before (French et al., 1960; Zipf, 1960) and is intended to control for the effects of individual differences in ability and motivation on the task prior to the reward instructions. It assumes that the faster *S* is going before motivating conditions are changed, the better able he will be to increase his speed subsequently. There is reason to accept this assumption, within limits. Several early studies on the effects of practice on performance of clerical tasks indicate that *S*s who are better at the beginning of practice improve more with practice than do the poorer *S*s (Anastasi, 1934; Thorndike, 1916).

It is instructive to compare the results obtained with this index of performance with several others that might have been used: absolute speed on the test trial $t_{3,1}$, and the difference between the test trial and the trial prior to change in motivation $t_{3,1}-t_{2,1}$. Using the same analysis as shown in Table 1 with these as dependent variables, there are no significant effects on $t_{3,1}$, and only Req is significant in its effect on $t_{3,1}-t_{2,1}$. Discounting the variable PrR for reasons given above, the ratio score obtains significance on two variables: the difference score on one variable and the absolute score on none. The ratio score would seem to be a more sensitive measure.

The use of the ratio is far easier than either matching groups or partialing out initial score with analysis of covariance, and possibly provides a better control. It is hoped that the index will lend itself to quantitative rather than merely directional predictions, which will apply to different tasks. Since the ratio would be seriously affected if *S*s were near an upper limit of speed, a measure of *S*'s upper limit of speed has been obtained in a more recent experiment. This measure will be used to clarify the type of index necessary to take account of differences in upper limit of performance as well as "operant level."

DISCUSSION

The importance of this study is seen as providing a simple method of studying and scoring human performance both so that comparisons may be made with animal studies of motivation and performance and so that other factors which are more relevant to human behavior may be considered. The findings of animal research on the variables of the present study were mentioned above. The majority of studies of human behavior utilizing these variables have examined choice rather than performance. We might expect, however, that direction and intensity of behavior might have some commonality and comparison with these findings is of interest.

We have tentatively concluded that *S*s increase their speed more when the probability of reward is 90% than when it is 50%. Edwards (1953) has shown that when postive expected value of a bet is held constant, preferences have a nonlinear relation to probability of winning and that 50% is preferred over 90%. When the expected value of bets differ, *S*s do not bet so as to maximize expected value (Edwards, 1954). Probability preferences have been related to various personality characteristics (Atkinson, 1957; Scodel, Ratoosh, & Minas, 1959). If preferences of this nature are operating, they would tend to confound the effect of PrR on performance. However, none of the choice studies seem to have considered very simple decisions where expected value was clearly different as would be comparable to the values in this study, for example: 50% chance of getting $1 versus a 90% chance of getting $1. The writer has some preliminary data to indicate that in such simple decisions, almost every person chooses the higher probability. Furthermore, Feather (1959b) has shown that *S*s wish for less probable goals but tend to choose more probable goals when the choice involves commitment to action. Another of his findings is compatible with a result of this study. In an achievement situation difficult tasks are preferred over easy ones while the reverse is true in a relaxed situation. The *S*s in the present study who were told that they had a small chance of succeeding were probably in a more extreme achievement situation than were those who were told that they were very likely to succeed, and perhaps placed a greater value on doing well.

Explicit comparisons of how these factors affect choice and performance must await research in which both choice among tasks and performance on these tasks are observed. Such research will be in order when more is understood about factors affecting simple choice and performance situations.

SUMMARY

Four variables were investigated for their effect on human performance. They were: amount of reward, probability of reward for successful performance, probability of success, and amount of increase in speed required. Probability of success and required increase in speed had significant effects on performance. No conclusion was possible for probability of reward with these data. Although amount of reward was not effective, a measure of desire to obtain the reward was significantly related to performance. A number of theories would suggest that the independent variables of this study would have multiplicative effects. However, no evidence of this was found.

Several different indices of performance were compared. A ratio of speed following the change in motivating conditions divided by speed before the change was shown to be more sensitive to the independent variables than either absolute speed or absolute change in speed.

REFERENCES

ANASTASI, A. Practice and variability. *Psychol. Monogr.*, 1934, **45**(5, Whole No. 204).

ATKINSON, J. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

CRESPI, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

CRUTCHFIELD, R. S. Determiners of energy expenditure in string pulling by the rat. *J. Psychol.*, 1939, **7**, 163–178.

EDWARDS, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

EDWARDS, W. Probability preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, **67**, 56–67.

EDWARDS, W. The prediction of decisions among bets. *J. exp. Psychol.*, 1955, **50**, 201–214.

FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, **66**, 150–164. (a)

FEATHER, N. T. Success probability and choice behavior. *J. exp. Psychol.*, 1959, **58**, 257–266. (b)

FRENCH, J. R. P., JR., MORRISON, H. W., & LEVINGER, G. Coercive power and forces affecting conformity. *J. abnorm. soc. Psychol.*, 1960, **61**, 93–101.

HULSE, S. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

LEWIN, K. The conceptual representation and the measurement of psychological forces. *Contr. psychol. Theor.*, 1938, **1**, No. 4.

LEWIS, D., & COTTON, J. W. Learning and performance as a function of drive strength during acquisition and extinction. *J. comp. physiol. Psychol.*, 1957, **50**, 189–194.

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

SCODEL, A., RATOOSH, P., & MINAS, J. S. Some personality correlates of decision making under conditions of risk. *Behav. Sci.*, 1959, **4**, 19–28.

THORNDIKE, E. L. Notes on practice, improvability, and the curve of work. *Amer. J. Psychol.*, 1916, **27**, 550–565.

TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–326.

ZIPF, S. G. Resistance and conformity under reward and punishment. *J. abnorm. soc. Psychol.*, 1960, **61**, 102–109.

(Received August 23, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 510–517

PERSEVERATION AS A FUNCTION OF DEGREE OF LEARNING AND PERCENTAGE OF REINFORCEMENT IN CARD SORTING[1]

ALBERT ERLEBACHER[2] AND E. JAMES ARCHER

University of Wisconsin

The purpose of this study was to investigate the effects of degree of learning and percentage of reinforcement, and especially their interaction, on perseverative errors in concept formation. Normally in the Wisconsin Card Sorting Test (WCST), *S* is required to sort a deck of cards according to the stimulus dimension *E* arbitrarily considers correct (Stage I). After *S* has achieved a criterion in Stage I, *E*, without informing *S*, changes the correct dimension (Stage II). If *S* erroneously continues to sort according to the formerly correct dimension in Stage II, the response is called a perseverative error.

Using the WCST, Grant and Berg (1948) and Grant and Cost (1954) had different groups of *S*s achieve different criteria before shifting to a new correct stimulus dimension. They found that there was an inverse relationship between the number of reinforced trials in Stage I and the number of perseverative errors in Stage II.[3] These findings seem to contradict the results of studies on simple learned responses in which a greater number of reinforcements usually leads to a persistence of the response. Grant and Cost (1954), however, believed that this inconsistency was only superficial. They suggested that *S* must learn to discriminate the change in the correct stimulus dimension and that this discrimination is facilitated by a large number of reinforcements.

This hypothesis is reminiscent of the hypothesis proposed by many researchers (e.g., Grant & Schipper, 1952; Humphreys, 1939) to account for increased resistance to extinction following partial reinforcement as opposed to 100% reinforcement in conditioning studies.

If Stage II is analogous to an extinction phase in simpler tasks we would predict that the number of perseverative errors should be inversely related to the percentage of reinforcement. Accordingly, three levels of percentage of reinforcement were investigated in the present study.

Furthermore, it seems very likely that the effect obtained under partial reinforcement is dependent on the number of trials during acquisition. It is not likely that there would be differential resistance to extinction under various percentages of reinforcement when only a few acquisition trials are given; if a larger number of trials were given however, there should be a differential resistance to extinction under various percentages of reinforcement. Perhaps this differential resistance to extinction even increases as a function of the number of trials. Indirect evidence for this hypothesis can be found in Grant and Berg (1948), and Grant and Cost (1954) who have shown that the number of reinforcements influences the number of perseverative errors. These findings, however, followed 100% reinforcement, and it is entirely possible that these results would not obtain under other reinforcement schedules. In fact, if a low percentage of reinforcement is similar to extinction, then with a greater number of trials, there should be greater generalization to extinction and thus, the greater the number of trials given in a low reinforcement first sorting stage, the longer should be the perseveration. Thus it seemed advisable that in addition to percentage of reinforcement, the number of trials during the first sorting stage should also be varied. The most interesting part of the results should be the Percentage Reinforcement × Degree of Learning interaction.

---

[1] This research was supported in part by NSF Grant G-1691 and while the first named author was a University Fellow.

[2] Now at Northwestern University.

[3] For similar results in animal learning studies see Reid (1953), Pubols (1956), and Capaldi and Stevenson (1957).

To increase the generality of the results the experiment involved replications with WCST concept varied. The WCST allows cards to be sorted according to form, number, and color. Thus, *S* was started in Stage I on one of the three dimensions and then switched to one of the other two.

METHOD

Experimental design.—Three variables were presented according to a 3 × 3 × 6 orthogonal design. These variables were percentage of reinforcement (25, 50, 100%), degree of first stage learning (10, 20, 40 trials), and stimulus dimensions in the two stages (form-number, form-color, number-form, number-color, color-form, color-number). In addition, one reinforcement sequence was used for half the *S*s in each partial-reinforcement group while another sequence was used for the other half.

Subjects.—Two *S*s were run in each of the 72 treatment conditions under partial reinforcement and 4 *S*s were run in each of the 18 treatment conditions under 100% reinforcement. This makes a total of 216 *S*s. The *S*s were all students in undergraduate psychology courses at the University of Wisconsin and were assigned to the various treatment cells at random with the restriction that the experiment be completed once before it was replicated. A time limit of 35 min. was arbitrarily chosen for learning in Stage I. Any *S* who exceeded this time limit was dropped from the experiment. Twenty-five *S*s were dropped for this reason; 12 were to have sorted according to color and were in the 25% group, 5 were in color-50%, 2 in color-100%, and 6 in form-25%. Twenty-three *S*s were dropped for reasons not related to any experimental conditions, mainly because of apparatus failure.

Materials and apparatus.—The materials used were a modification of the WCST. The WCST consists of cards showing figures which vary in three stimulus dimensions: form (circles, triangles, crosses, and stars), number (1, 2, 3, and 4), and color (red, green, yellow, and blue). For example, one card might show four red crosses and another might have one green circle.

The WCST cards are normally sorted into four piles, the first labeled by one red triangle, the second by two green stars, the third by three yellow crosses, and the fourth by four blue circles. Thus, *S* can make one of four responses. He can sort according to form, number, or color, or he can make a response which is none of these and which is unanalyzable.

In order for each response to be unambiguously classified only 24 of the standard 64 cards can be used. For example, none may have a combination of red and one, or red and triangle, or one and triangle. Cards with combinations of levels for the other label cards were also eliminated. The set of 24 cards used in this experiment thus represented all possible combinations of the four forms, four numbers, and four colors which did not represent any of the above combinations.

The cards were not actually presented to *S*s, but were projected on the back of a milk-glass screen by a Dunning Animatic 16 mm. strip-film projector. The *S* sat in a 4 × 4 × 8 ft. booth directly in front of the screen, which was approximately 11½ in. wide and 7½ in. high. The pictures were approximately 7¼ in. square.

The cards used were those of the "unsystematic" type (Grant & Curran, 1952). That is, the position of figures on a card was unrelated to the number of figures. The film strip was a closed loop and contained 192 pictures. Each combination of dimensions appeared eight times but unlike the standard WCST, four different configurations were used for each combination of dimensions.

A special procedure was followed in ordering the presentation of the cards. Essentially the entire sequence represented eight different orders of the 24 cards. Four sets of the 24 cards were made up, the sets differing only in card configuration. The 24 cards of one configuration set were randomly ordered. The cards of the other three sets were ordered systematically on the basis of this first order. The four sets of cards followed each other consecutively making up the first 96 cards. A different order for the second 96 cards was similarly obtained. A detailed account of the ordering procedure is presented by Erlebacher (1960).

Four response-label cards were placed under a clear plastic sheet across the center of *S*'s control box. The response cards from left to right were: one red triangle, two green stars, three yellow crosses, and four blue circles. At the rear center of the box was a white light; at the right rear of the box was a red light. In front of each response card was a button which *S* pressed to make his response. The white light was the reinforcement light and the red light indicated to *S* when the experiment was over. When a new picture appeared on the screen, *S* pressed one of the four buttons, taking as much time as he pleased. After he pressed the button, he waited until a new picture came into the window. If his response were reinforced, the white light came on at the moment of his pressing the button and it stayed on for approximately 2.25 sec. The next picture came into the screen approximately 2.60 sec. after the button was pressed.

The two partial reinforcement schedules were 25% and 50%. The percentage refers to the percentage of times *S* was told he was correct (the white light went on) *when he actually was correct*. Thus, whether any correct response was reinforced was contingent on *S*'s previous number of correct responses.

A schedule of reinforcement was made up which took into account 200 trials of correct responses on the part of *S*. These 200 trials were divided into blocks of 20 trials such that the percentage indicated was maintained in each block. The sequence of reinforcement-nonreinforcement was random with the restrictions that the first trial of each block be reinforced and that at the most, eight positive reinforcements could occur in a row. In order for the results of the experiment to be generalizable to various reinforcement sequences, half of the *S*s were started at the beginning of this schedule of 200 trials and the other half were started at the middle of the schedule. This difference in starting position represented the two reinforcement sequences mentioned earlier. The reinforcement schedule was repeated if *S* required more than 200 trials. All timing, reinforcement, etc., was controlled by automatic apparatus, consisting mostly of relay circuits, which was kept in a remote room away from *S*, so that *S* could not hear the apparatus in operation.

Procedure.—After *S* made either 10, 20, or 40 consecutively correct responses in Stage I the "correct" dimension for sorting was changed. Except for the change in reinforcement *S* was given no information about this change. He was now reinforced for every correct sorting on one of the other two stimulus dimensions and continued on this second stage until he had made 10 consecutively correct responses. If *S* did not reach this latter criterion at the end of 300 trials on the new dimension, the experiment was terminated.

Instructions.—The *S* was told that for each of the pictures appearing on the screen he was to press one of the four buttons in front of him and that only one of the four buttons would be correct. If he were correct, the white light might go on, but it would not necessarily do so. If he were incorrect, however, the white light would not go on. (It was necessary to give *S*s partial-reinforcement instructions, i.e., tell them that the white light would not always go on even if they were correct, because a preliminary pilot study indicated that most *S*s could not reach even a moderate criterion—especially in the 25% reinforcement group—if these instructions were not given.) All *S*s, including those of the 100% group, were given partial-reinforcement instructions. No mention was ever made of stimulus dimensions, and if *S* asked a question referring to such dimensions, he was not given a direct answer involving such a reference. No instructions about a change in the correct stimulus dimension were given.

RESULTS

The responses during Stage I (learning) and Stage II (extinction) were analyzed separately. In Stage I, the responses were analyzed in three categories: number of correct responses, number of errors, and total number of responses. These measures do *not* include the criterional trials, since these were different for various groups. In Stage II, the analyses were on number of perseverative errors, number of nonperseverative errors, and total number of responses. The criterional trials were included in Stage II measures. The first response of Stage II was considered a perseverative error since this would not differentially affect the scores of the various groups.

Stage I.—The total variance for Stage I was analyzed and showed that percentage of reinforcement was a significant variable affecting all three response measures ($P < .001$). Figure 1 presents the mean values for all three measures for all three reinforcement groups. As the percentage of reinforcement increases, all three scores decrease. It must be noted that the values for the 25% group are somewhat underestimated since the scores of those *S*s who did not reach the criterion (most of whom were in the 25% group) are not included.

The degree of learning was also a significant variable for total correct responses. The means were 20.0, 16.2, and 34.0 for the 10, 20, and 40-trial groups, respectively ($F = 4.83$, $df = 2/171$, $P < .01$).

Another significant term in the analysis was the Percentage Reinforcement × Dimension interaction for total responses ($F = 2.44$, $df = 4/171$, $P < .05$). This relationship is given in Fig. 2, which shows that the large factor contributing to this interaction is that sorting according to number was more difficult at 50% reinforcement than was sorting according to form, while on 25% reinforcement the order was reversed. Again it must be pointed out that the figure for total responses when sorting according to color is underestimated especially under 25% reinforcement, as is also the figure for form on 25% reinforcement.[4]

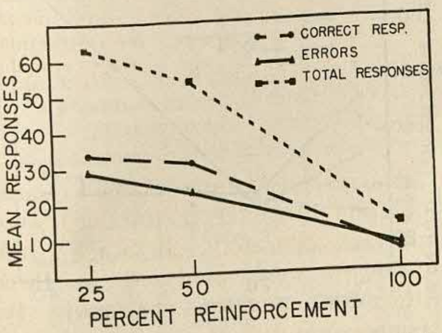


FIG. 1. Mean total responses, mean total correct responses, and mean total errors in Stage I as a function of the percentage of reinforcement.

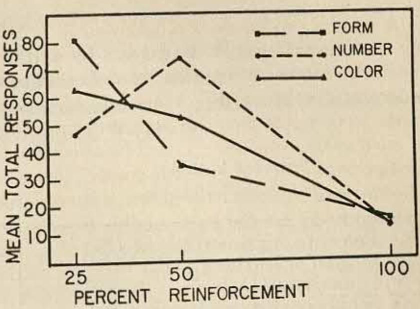


FIG. 2. Mean total responses in Stage I as a function of the percentage of reinforcement and the correct stimulus dimension.

Stage II.—Table 1 summarizes the analyses of variance of the three response measures obtained during Stage II.[5] The term *dimension* in

---

[4] Several significant interaction terms in the analysis involve sequence within percentage reinforcement which seems to indicate that the sequence used is effective according to what conditions it is used with. No estimation analysis is given for these interactions or for any terms in this paper involving sequences since sequences were not systematically varied and thus no conclusions could come from such an analysis. The reason for having introduced various sequences was only as a check for the generality of the results and the analyses of variance seem to indicate that any future studies on partial reinforcement, in this area, should involve several sequences of reinforcement.

[5] Since the experiment was terminated at the end of 300 responses, in order to obtain

TABLE 1

ANALYSES OF VARIANCE OF TOTAL RESPONSES, PERSEVERATIVE ERRORS, AND NONPERSEVERATIVE ERRORS IN STAGE II

Source of Variation	*df*	*F* Ratios		
		Total Responses	Perseverative Errors	Nonperseverative Errors
Percentage reinforcement (%)	2	1.90	2.90	.82
Degree of learning (DL)	2	1.98	3.49*	.64
Dimension (D)	5	1.87	.77	4.74**
Sequence/Percentage reinforcement (S/%)	2	.14	.33	1.50
% × DL	4	3.33*	3.44*	1.96
% × D	10	.66	.67	1.06
DL × D	10	1.53	1.51	1.61
DL × S/%	4	.20	.23	.33
D × S/%	10	1.10	.74	1.01
% × DL × D	20	.94	.84	.66
DL × D × S/%	20	1.20	1.02	.67
Error (*MS*)	126	(6986)	(5388)	(720)
Total	215			

* $P = .05$.
** $P = .01$.

Table 1 refers to a particular dimension sequence such as color-form, i.e., *S* had to sort according to color in Stage I and according to form in Stage II.

The analyses show that the degree of Stage I learning is a significant variable in determining the number of perseverative errors. The mean total perseverative errors were 49.9, 35.99, and 68.19 for the 10, 20, and 40-trial criterion groups, respectively.

The Percentage Reinforcement × Degree of Learning interaction is also a significant term in the analysis of variance for both total responses and total perseverative errors. This relationship is shown for perseverative errors in Fig. 3. It should be noted that the groups do not rank 100%, 50%, 25% going from fewest to most perseverative errors for *all* degrees of original learning. Such an ordering would have been expected if the normally found partial-reinforcement effect were so strong that it overshadowed any effect due to other

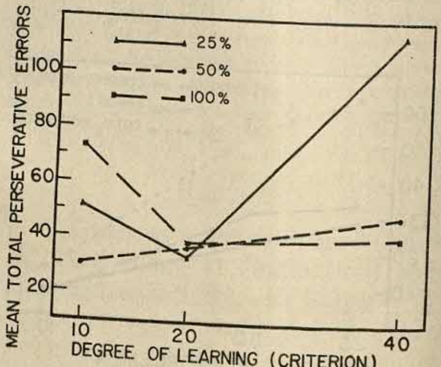


FIG. 3. Mean total perseverative errors in Stage II as a function of the percentage of reinforcement and the degree of learning in Stage I.

---

unbiased results for those *S*s who had not reached criterion, the scores were changed to percentage out of 300. This kind of a score normally requires an arc sin transformation, so a second analysis was performed on the transformed scores. The entire transformation consisted of dividing each score by 300, taking the square root of that result, and then finding the arc sin of that square root. This second analysis gave essentially the same results as are shown in Table 2. Since no different conclusions are reached when transformed scores are used, the results given below will be only for untransformed scores.

TABLE 2

Number of *S*s Making 300 Responses
Stage II

Degree of Learning (Trials)	% Reinforcement			Total
	25	50	100	
10	2	0	4	6
20	0	2	1	3
40	2	2	2	6
Total	4	4	7	15

variables. Also to be noted is the fact that the difference between the 25%, 50%, and 100% reinforcement groups does not increase as the number of criterional trials increases.

Fifteen *S*s made 300 responses, at which time the experiment was terminated. Table 2 shows in which condition these *S*s served.

The mean total perseverative errors were 66.3, 36.9, and 50.9 for the 25%, 50%, and 100% reinforcement groups. These differences were not significant.

The interaction for total responses is similar to that for perseverative errors. Since there is little difference in the shape of the curves for total responses and those shown in Fig. 3, separate figures are not presented here.

In addition to these variables, it is shown in the analyses of variance table that there is a significant effect due to the dimension sequence but only as far as nonperseverative errors are concerned. This relationship is shown in Table 3.

TABLE 3

Mean Total Nonperseverative Errors as a Function of the Correct Stimulus Dimensions in Both Stages

Correct Dimensions		Mean Total Nonperseverative Errors
Stage I	Stage II	
Form	Number	5.56
Form	Color	8.83
Number	Form	3.06
Number	Color	29.72
Color	Form	5.78
Color	Number	10.61

Discussion

Stage I.—As was seen, performance in Stage I was affected by the percentage of reinforcement. This result was not unexpected. As the percentage of reinforcement increased, both the number of errors and number of correct responses decreased. Though it may seem reasonable that errors should decrease and therefore paradoxical that correct responses decrease also, the apparent inconsistency is due to requiring *S* to achieve a criterion.

It is also to be expected that the more stringent the criterion, the longer it takes to reach it. This relationship held for the highest criterion (40 consecutively correct responses). No difference, however, was seen between the two other criteria (10 or 20 consecutively correct responses).[6] The three criterion groups need not have differed in number of errors. Because the criterion was one of 10, 20, or 40 consecutively correct responses, one or two errors between the attainment of a lower criterion and the attainment of the final criterion could greatly increase the number of correct responses made, but have little effect on the error score. Such slight differences in number of errors could not show up reliably in this experiment because of the relatively large variability among *S*s.

Stage II.—It was seen that the degree of learning in Stage I resulted in a differential number of perseverative errors in Stage II. Unlike previously cited studies, the groups with many overlearning trials (40 consecutively correct

[6] Although the number of correct responses required to reach criterion by the 20-trial criterion group was slightly less than by the 10-trial group, this difference can only be attributed to chance since it is logically impossible for a 10-trial criterion to be more difficult than a 20-trial criterion.

responses) made *more* perseverative errors than did the groups with fewer criterional trials. The 20-trial criterion group made fewer perseverative errors than the 10-trial criterion group. Little can be said about these results since there was a significant Degree of Learning × Percentage Reinforcement interaction and it is there that we must look for any answers. The same can be said for percentage reinforcement. Although 25% reinforcement resulted in a greater number of perseverative errors, the interaction is the important determinant of the results.

Under 100% reinforcement, the results were clearly the same as those of other studies cited in the introduction: in general, the fewer the criterional trials, the greater the number of perseverative errors. In the study by Grant and Cost (1954) it was also shown that there was little difference between the 20-trial criterion group and the 40-trial group. This same result was shown here.

However, under 50% reinforcement the opposite results were obtained. Although the effect is slight, the greatest number of perseverative errors was made by the groups required to reach a 40-trial criterion, the second most frequent number of perseverative errors was made by the 20-trial group, and the least number of perseverative errors was made by the group required to reach a 10-trial criterion. This result could be explained in terms of the hypothesis stated in the introduction. That is, if the reinforcement pattern in Stage I is similar to that in Stage II, the change would be difficult to detect and more perseverative errors would be made. Furthermore, under partial reinforcement, a greater number of trials in Stage I would increase the length of runs of nonreinforced trials which would be learned by *S* not to be infirmatory in nature. Thus the pattern of reinforcement for the same response in the second stage (0%) would be more similar to the first stage pattern for those groups with greater numbers of criterional trials. This, then, results in slower extinction, or a greater number of perseverative errors.

Somewhat similar results as in the 50% groups were obtained for 25% reinforcement. After 40 criterional trials there was clearly a larger number of perseverative errors than after the other criteria. This fits into our analysis. However, the 10-trial criterion group made more perseverative errors than the 20-trial criterion group. No explanation could be found for this latter result.

Although predictions were made for the effect of various criteria within the three percentage reinforcement groups, nothing was said of the relative position of all nine groups with respect to each other. If it were supposed that the effect of the percentage of reinforcement on extinction, as it is normally found, were so strong that it applied under all conditions of degree of learning, it might have been predicted that the groups—no matter what the degree of learning was—would rank 100%, 50%, 25% going from fewest to most perseverative errors. This was clearly not the case. In addition to this possible prediction, it might have been expected that the difference between the various percentages of reinforcement would increase, the greater the number of criterional trials. This again was not found.

Two methodological factors in the present experiment may account for the unexpected results. One of these factors was that all *S*s received the same instructions. Thus, even the 100% groups received partial reinforcement instructions. These instructions were given so that the only differences found in the experiment would be due to the experimental conditions alone and not due to differences in instructions. For 100% reinforcement, though, these instructions may have caused *S*s to make more perseverative errors than they would normally tend to make. Unfortunately the values in this experiment cannot be directly compared to those found by Grant and Cost (1954) since they used the full WCST deck of cards and only 24 of these were used in this experiment.

The other methodological factor was

the impossibility of determining in advance when the last criterional response would be made by any *S* in the partial reinforcement groups. Therefore it was impossible to reinforce the last criterional trial. Since this was the case, there was no definite demarcation between Stage I and Stage II in the partial reinforcement groups. For these groups, especially for the 25% groups, extinction therefore actually started before the last criterional trial was reached. Thus the curves for the 50% groups and especially for the 25% groups are depressed, and more perseverative errors were made than is shown by the results.

It was shown that *S*s who sorted according to color in Stage II made more nonperseverative errors than *S*s who sorted according to form in Stage II, when both groups sorted according to number in Stage I. This result is in agreement with the results of Grant and Curran (1952) which show that at Stage II more nonperseverative errors are made when sorting under color than when sorting under form.

SUMMARY

An experiment relating perseveration to percentage of reinforcement and degree of learning was performed using a modified form of the Wisconsin Card Sorting Test (WCST). The results showed that there was a significant interaction between these two variables when perseverative errors were considered. Under 100% reinforcement, the greater the degree of learning the less the perseveration (within limits). Under partial reinforcement, somewhat of an opposite result was found indicating that the greater the degree of learning, the greater the number of perseverative errors. These results were discussed in terms of the similarity of the pattern of reinforcement during the acquisition stage and the nonreinforcement of the same response during the extinction stage. Supposedly, similarity leads to generalization and thus, slower extinction (greater perseveration).

The learning data showed that the more stringent the criterion, the more difficult the solution of the problem was, though this held true only for the most stringent criterion. Percentage reinforcement also affected the learning time in that the greater the percentage of reinforcement the simpler was the task.

REFERENCES

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

ERLEBACHER, A. Perseveration as a function of degree of learning and percentage reinforcement in card sorting. Unpublished master's thesis, University of Wisconsin, 1960.

GRANT, D. A., & BERG, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *J. exp. Psychol.*, 1948, **38**, 404–411.

GRANT, D. A., & COST, J. R. Continuities and discontinuities in conceptual behavior in a card sorting problem. *J. gen. Psychol.*, 1954, **50**, 237–244.

GRANT, D. A., & CURRAN, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.

GRANT, D. A., & SCHIPPER, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

PUBOLS, B. H. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

(Received August 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 518–522

SHORT-TERM RETENTION AS A FUNCTION OF AVERAGE STORAGE LOAD AND AVERAGE LOAD REDUCTION[1]

LYNE STARLING REID, KENNETH E. LLOYD,[2] H. RAY BRACKETT, AND WILLIAM F. HAWKINS[3]

University of Virginia

A series of experiments (Lloyd, Reid, & Feallock, 1960) has been previously reported in which short-term retention was studied in a task that required *S*s to recall sequentially presented English words. The task differed from the one used to study classic memory span in that the presentation of items and their recall were not discretely separated. Instead, both occurred in a variety of serial orders over an extended period of time. Each item to be remembered was followed at some point in the sequence by a cue which specified that *S* must recall the item. The recall point may have followed immediately the presentation of the item, or may have been separated from it by several other items, by other recall points, or by both. Throughout the task *S* was presented with new words he had to remember and from time to time was required to recall words previously presented. Recall errors in the task were found to be related in a systematic manner to the average number of items *S* was required to remember. Furthermore, if the items are familiar, *S* appears to be equally capable of storing items from different sized populations so long as the average number of items to be remembered, i.e., the average storage load, remains constant.

In these earlier studies *S* recalled only one item at each recall point. In the studies to be reported the generality of average storage load is assessed for sequences in which *S* is required to recall more than one stored item at a recall point. As in the original experiments the words to be remembered are instances of a word class and the class names serve as recall points. The following series of words is an example of such a sequence: Berlin, Willow, Moscow, Dartmouth, Elm, CITY, Golf, Ford, Soccer, TREE, COLLEGE, ... When the word CITY occurs *S* should respond with "Berlin" and "Moscow." Similarly, at TREE the correct recalls are "willow" and "elm." At COLLEGE only one item, "Dartmouth," is appropriate.

The above sequence can be used to make clear the concept of average storage load and a new summarizing measure *average load reduction*. At the recall point CITY, *S* has been presented five items, and, thus, has a storage load of 5. Upon the recall of "Berlin" and "Moscow" this is reduced by 2, but before TREE occurs *S* is presented three further items increasing the load to 6. Again at TREE the load reduction is 2, so to this point average storage load is 5.5 and average load reduction 2.0. In the present experiments these averages were computed over much longer runs, but were obtained in like manner by simply averaging the load and reduction values at each of the recall points in a sequence.

Two experiments are reported. The first determines the effect of various values of average storage load upon the recall of sequentially presented items when average load reduction is held constant, but at a value greater than 1. In Exp. II average load reduction is varied while average storage load is held constant.

---

[1] The reported studies were supported by Contract No. DA-49-007-MD-537 between the University of Virginia and the Office of The Surgeon General, Department of the Army for which Richard H. Henneman is principal investigator.

[2] Now at Washington State University.

[3] Now at George Peabody College.

Experiment I

Method.—The design of the experiment parallels that of Exp. I in Lloyd et al. (1960). The *S*s stored and recalled words presented to them aurally by a magnetic tape recorder. In an experimental session they listened to two sequences. Each sequence was composed of 12 subsequences constructed to have any one of four values of average storage load: 2.5, 3.5, 4.5, and 5.5. Within a sequence each value of average storage load appeared three times in a partially counterbalanced order.

Average load reduction in all sequences was 2. This was accomplished by using individual reductions of either one, two, or three stored items at each recall point. A reduction of 2 was the most frequent value.

Throughout a sequence the values of the storage loads and load reductions at individual recall points occurred in a variety of combinations with the exception of the 2.5 subsequences. When average load and average reduction are nearly equal as is the case in the 2.5 subsequences, the individual loads and reductions must also be nearly equal. For this reason, in the 2.5 subsequences the load and reduction were equal at one-half the recall points. At these points *S* had to recall all the items being stored. Since it is impossible to avoid a reduction of load to 0 in the 2.5 subsequences if an average load reduction of 2 is to be maintained, reductions to 0 were introduced into the subsequences of higher storage loads even though they are unnecessary.

All subsequences contained 20 items which *S* had to remember and for which he was scored. With an average load reduction of 2, this means there were 10 recall points for these items. It was necessary, however, to insert extra items in order to maintain the desired loads at recall points near the ends of the subsequences. These extra items were called for immediately after the tenth recall point in the subsequence and were not scored.

The items were instances of classes of English words, and the class names were used as recall points. The eight class names were AUTOMOBILE, TREE, CITY, METAL, SPORT, CLOTHING, COLLEGE, and BIRD. Each class contained 8 items judged to be familiar to *S*s. However, since the words were probably not equally familiar, the *S*s were given preliminary familiarization with the recall material at the beginning of the first experimental session. They were presented with a list of the 64 items grouped into the eight classes and were allowed 5 min. to become familiar with the items. In addition, they were permitted to keep the list for reference during the first experimental session and prior to each subsequent session were required to study the list for 3 min.

Four sequences were recorded on the magnetic tape. The sequences differed in the serial order of the 12 subsequences, and in the particular words that were substituted at the different loads. A presentation rate of one item or recall point every 3 sec. was used.

The 12 *S*s, University of Virginia undergraduates, were instructed to remember each item as they heard it on the tape and to recall the items when the appropriate class name was heard. It was pointed out that one, two, or three words might be required at any recall point, and that any substitution, omission, or addition would be scored as an error. They were encouraged to guess if uncertain about a recall. Each *S* responded to two sequences a day for 4 days.

TABLE 1

Mean Errors as a Function of Average Storage Load during the 4 Days of Practice

Day	Average Storage Load							
	2.5		3.5		4.5		5.5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	1.94	0.95	4.05	1.40	6.76	1.25	8.33	1.59
2	0.57	0.41	1.65	1.21	5.22	1.98	6.47	1.57
3	0.44	0.41	0.83	0.74	3.97	1.31	5.04	1.34
4	0.10	0.05	0.47	0.44	3.01	1.16	4.93	1.29

Results.—All *S*s on each of the 4 days responded to six subsequences at each value of average storage load. The means of errors over these repeated measures are presented in Table 1. As in the previous studies (Lloyd et al., 1960) recall errors increase systematically with increases in average storage load. For the analysis of variance over the 4 days of practice a square-root transformation of the data was made since the errors for low storage loads are skewed. From this analysis storage load, practice, and the Load × Practice interaction are found to be significant beyond the 1% level.

Since the recall scores on Day 4 are considered the best estimate of stable performance a second analysis was performed. Table 2 summarizes the trend analysis of transformed mean error scores for each *S* on this day. The overall trend is highly significant. Although the first three components are significant, 95% of the overall-trend sum of squares can be attributed to the linear component. The significant quadratic and cubic components probably arise from the relatively small error differences at the lower storage loads compared to the much larger error difference beyond the 3.5 load.

TABLE 2

TREND ANALYSIS OF TRANSFORMED SCORES OF AVERAGE STORAGE LOAD

Source	*df*	*MS*	*F*
Overall trend	3	6.42	214.00*
Linear	1	18.29	457.25*
Quadratic	1	0.22	11.00*
Cubic	1	0.75	25.00*
Between individual means	11	0.16	5.33*
Between individual trends	33	0.03	
Linear	11	0.04	
Quadratic	11	0.02	
Cubic	11	0.03	
Total	47		

* *P* <.01.

EXPERIMENT II

Method.—Values of average load reduction are obviously dependent upon values of average storage load; there can never be greater reduction than there is load. It was for this reason in Exp. I where storage load was varied that load reduction was held constant. Storage load was in Exp. II held constant at 4.5, while average load reduction was varied through six values: 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5.

Two subsequences were constructed for each value of load reduction. The 12 subsequences were similar to those of Exp. I; their length, however, was increased to 35 items in order to provide adequate sampling of high-reduction values. Furthermore, the number of recall points varied from subsequence to subsequence. This was necessarily the case since the number of recall points in a subsequence decreases as average load reduction is increased. There were 35 recall points in the 1.0 subsequences, 23 at 1.5, 18 at 2.0, 14 at 2.5, 12 at 3.0, and 10 at 3.5.

Four sequences were prepared using the items and class names of Exp. I. The sequences differed in the serial order of the six subsequences and in the words assigned to the subsequences. Presentation rate for the items was one every 3 sec., and the duration of each recall point was 4 sec.

Instructions to *S*s were similar to those of Exp. I. Ten University of Virginia undergraduates responded to two sequences a day during 4 days of practice.

Results.—On each of the 4 days all *S*s responded to two subsequences at each value of average load reduction. The mean errors on these repeated measures were averaged for each *S* at each value of average reduction and are shown in Table 3. It is apparent that when average storage load is held constant recall errors systematically decline as average load reduction is increased. An analysis of variance over the 4 days of practice reveals load reduction

TABLE 3

MEAN ERRORS FOR AVERAGE LOAD REDUCTION DURING THE 4 DAYS OF PRACTICE

Day	Average Load Reduction											
	1.0		1.5		2.0		2.5		3.0		3.5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	21.70	2.13	14.65	2.65	12.75	3.56	9.30	3.17	7.75	3.61	5.65	3.13
2	17.55	3.46	13.50	3.39	8.90	2.95	7.15	3.01	5.35	3.32	4.00	3.42
3	15.85	2.73	10.30	3.83	6.40	2.46	4.95	2.52	3.95	2.67	2.30	2.00
4	14.75	3.54	9.90	3.56	5.90	3.44	4.85	2.71	3.65	2.99	2.30	2.13

and practice to be significant beyond the 1% level; the Reduction × Practice interaction was not significant.

Table 4 contains a trend analysis of the data on Day 4. Since the errors for the higher values of load reduction are skewed a square-root transformation of the data was made. The overall trend, and both the linear and quadratic components are significant beyond the 1% level. Again, as in Exp. I, most of the variance in the overall trend is accounted for by the linear component.

TABLE 4

TREND ANALYSIS OF TRANSFORMED SCORES OF AVERAGE LOAD REDUCTION

Source	*df*	*MS*	*F*
Overall trend	5	8.20	82.00*
Linear	1	39.21	261.40*
Quadratic	1	1.43	20.43*
Cubic	1	0.17	2.83
Quartic	1	0.08	—
Quintic	1	0.13	—
Between individual means	9	2.50	25.00*
Between individual trends	45	0.10	
Linear	9	0.15	
Quadratic	9	0.07	
Cubic	9	0.06	
Quartic	9	0.10	
Quintic	9	0.14	
Total	59		

* $P < .01$.

DISCUSSION

In these studies *S* stored and recalled words in a single sequence for a period of 15 to 20 min. During this time the words and their recall points occur in a variety of combinations which introduce many variable conditions that can affect performance. For example, the many different patterns of presentation and recall produce variations which have, by analogy to long term retention paradigms, proactive and retroactive effects. In this regard, however, there are two important differences between the way these conditions vary in the present studies and the manner they typically vary in a long term retention study. First, in the long term retention studies a relatively few values of one or a small number of variables are manipulated in a systematic fashion while the attempt is made to hold all other conditions constant. In the two studies under consideration, on the other hand, many values of a number of variables are permitted to vary in a nonsystematic fashion. Second, in long term retention studies conditions usually can be varied in an independent manner; within the sequences used in the present studies, however, prior and interpolated items and recall points cannot be manipulated independently of one another. A given pattern of items and recall points restricts to a considerable degree the previous and later occurrence of other items and recall points.

Both these factors make highly desirable variables which summarize the variety of conditions and dependencies within the task. The earlier experiments of Lloyd et al. (1960) established the fact that, in tasks requiring the recall of only one item at each recall point, the average number of items *S*s are required to remember is related in a reliable manner to their recall errors. Experiment I extends the generality of these findings to situations in which the recall of more than one item at a recall point may be required. Furthermore, the highly significant and consistent differences found between the different values of average storage load point up the sensitivity of performance to this variable. As was pointed out above, *S*s are storing and recalling words in a sequence for a period of 15 to 20 min. During this period load changes occurred approximately every 1.5 min., and recall paralleled these changes throughout the sequence.

Experiment II demonstrates that a second important summarizing variable is significantly related to performance in such tasks. At first it may seem surprising that improved retention is related to increases in average load reduction, but a consideration of the variables contributing to its effect makes such an outcome reasonable. Two important changes took place in Exp. II with increases in load reduction. The first, a decrease in number of recall points, is simply a correlated condition of the sequential task. This decrease in recall points does, however, reduce the average amount of interpolated activity between the presentation of items and their recall, and may account, in part, for the improved performance. The second, the fact that increased load reduction means more items from a single class are being stored, provided the opportunity for the grouping of related items. In this latter respect average load reduction is obviously related to the phenomenon of clustering (Bousfield, 1953) and to Miller's (1956) concept of the chunk.

Although the reduction in recall points and the opportunity for grouping are in a sense confounded in the present study, they can be varied independently. If, for example, arbitrary labels are used to designate the categories into which words are grouped for recall, the extent to which the words grouped together are related can be varied over a wide range of values. The pattern of their presentation and recall, on the other hand, can be held constant by merely substituting the different words into the same sequences. An investigation of the influence of relationship among grouped items upon performance is under way. Such an investigation points up one of the major advantages of the present method, i.e., a number of complex conditions can be held constant while one or more experimental variables are systematically manipulated.

SUMMARY

Two experiments are reported in which short-term retention was studied in a situation requiring *S* to remember familiar English words until their recall was requested. During the time *S* was required to remember these items he was presented other words or asked to recall previously presented items. Experiment I demonstrated that in a task where the recall of more than one item could be requested at a given point in the sequence, errors increased systematically as the average number of words to be remembered over blocks of recall points, average storage load, was increased. In Exp. II it was found that when the average number of items to be remembered was held constant, errors decreased as the average number of items to be recalled at given recall points, termed average load reduction, was increased.

REFERENCES

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

Lloyd, K. E., Reid, L. S., & Feallock, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

(Received August 26, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 523–531

THE EFFECTS OF INDUCED MUSCLE TENSION DURING TRACKING ON LEVEL OF ACTIVATION AND ON PERFORMANCE[1]

LAWRENCE R. PINNEO[2]

Allan Memorial Institute, McGill University

Since Courts' (1942) review, several experiments have been reported on the effects of induced muscle tension on performance, but to the author's knowledge none have been concerned with the basis of the relationship. In a theoretical paper Meyer (1953) suggested that the effects of induced muscle tension on performance are the result of simultaneous responses occurring at the level of the motor cortex. Though a useful concept, Meyer fails to consider the involvement of the reticular activating system and level of arousal in muscle tension phenomena (Delafresnaye, Adrian, Bremer, & Jasper, 1954).

In contrast to Meyer (1953), Malmo (1959, p. 370) suggests that induced tension may be one of the ways in which activation level can be varied, and that activation level, in turn, is a function of the reticular activating system. If induced tension is indeed a reliable means of varying activation level, then in addition to behavioral effects, induced tension should also produce regular and consistent changes in the various physiological indicants of activation. Experiments by Freeman and Simpson (1938) encourage this idea. They have shown that palmar conductance increases as a function of muscle tension induced in the legs, and Freeman (1938) has shown that palmar resistance changes linearly with induced muscle tension during performance. However, these experiments were unfortunately limited to the one physiological measure of palmar conductance, while Malmo (1959) and Schnore (1959) have shown the need for a wide coverage of physiological functions in studies of activation. The purpose of this experiment was the exploration of the relation between induced muscle tension, activation, and level of performance.

METHOD

Subjects

The *S*s were 38 McGill University undergraduate and graduate males, ranging in age from 18 to 30 yr. All were paid $5.00 for the 3-hr. session. Only right-handed *S*s with normal hearing were used.

---

[1] This paper is part of a thesis presented to the Faculty of Graduate Studies of McGill University for a PhD degree. It was supported in part by the Medical Research and Development Division, Office of The Surgeon General, Department of the United States Army: Contract Number DA-49-007-MD-626; Defence Research Board, Department of National Defence, Canada: Grant Number 9425-04; National Institute of Mental Health, National Institutes of Health, United States Public Health Service: Grant Number M-1475; National Research Council of Canada: Grant Number A.P. 29; and Department of National Health and Welfare (Canada): Grant Number 604-5-69.

The author expresses his thanks to Robert B. Malmo for his continuing interest, guidance, and support; D. J. Ehrlich, P. F. MacNeilage, and L. D. Rust for their discussion and criticism of the experiment; W. Mundl and J. F. Davis for assistance with the apparatus; and Jeanne M. Pinneo for her assistance in carrying out the experiment.

[2] Now at the Institute of Physiology, University of Pisa, Italy.

Apparatus[3]

Muscle tension induction system.—A modified hand dynamometer was used for inducing tension. When the spring steel of the dynamometer was compressed, two strain gauges mounted on each side of the spring were deformed. This resulted in a change of the electrical resistance of the wire in the strain gauge proportional to the strain or the grip on the dynamometer. Each strain gauge was an arm of a balanced Wheatstone bridge. Changes in the resistance of either or both of the strain gauges upset the electrical balance of the bridge. The change in signal due to this imbalance was fed into a push-pull, strain gauge amplifier.

Part of the amplified signal was used to continuously monitor and permanently record *S*'s instantaneous performance, and part was fed back to *S* to enable him to maintain his level of pull within prescribed limits. These limits were set by an ac microammeter relay equipped with two movable contacts. Both contacts could be positioned anywhere on the meter face and their relative positions could also be varied. By calibrating the meter deflection in kilograms, and using each of the meter-relay contacts as a switch to an audio oscillator, "low" and "high" tension limits could be determined. When *S* did not squeeze the dynamometer sufficiently hard, he heard a 600-cycle tone via one contact, while if he squeezed too hard he heard an 800-cycle tone via the other contact. The desired pressure was obtained when *S* gripped the dynamometer such that he heard no tone at all.

Auditory tracking system.—This system was similar to that described by Davis, Stennett, and Quilter (1957), except that tracking was done with the right foot instead of with the hand. In addition the function generator used in this system consisted of a potentiometer bridge circuit whose imbalance signified error in direction and intensity. The *S*, who was blindfolded for purposes of EEG recording, sat in a semireclining chair and performed a simple tracking task in the following manner. He first depressed the foot pedal until it came against a mechanical stop after a movement of 8° of arc. He next allowed the pedal to return to its normal position at the same rate at which the pedal was depressed. There was a stop at the upper, as well as at the lower terminal position of the pedal, and in order to entirely avoid error it was necessary for the foot to move off each stop immediately. In tracking, a constant rate of movement was required. If the rate of pedal movement down and up was in exact conformance with that of the function generator, *S* was "on target" and he heard no sound in his earphones. Failure to keep pace with the function generator was signaled to *S* by means of a 1000-cycle tone in his earphones. The loudness of this tone varied as a function of *S*'s "distance off target," and direction of error was cued by whether the tone appeared in the right or the left earphone. "Down" errors (i.e., pressing down on the pedal at too fast a rate or lagging behind on the way up) were indicated by the tone appearing in the earphone on the right; conversely, "up" errors were indicated by the tone appearing in the earphone on the left.

The number of up and down errors made per trial and the total distance *S* was off target (DOT) were the measures of performance. Number of errors were recorded on two electrical impulse counters, and DOT was recorded graphically by means of an electronic integrator with a 4-sec. discharge rate (Davis, 1959). This latter system was calibrated such that a certain distance off target in degrees produced a given deflection on the ink writer, which was then measured with special scales.

Physiological recording apparatus.—Three chart drives with ink-writing galvanometers were used for recording physiological tracings and error data. The primary record for electromyograms (EMG) from active and passive muscles, two EEG measures, and the heart rate (EKG), were recorded on a standard Grass IVA electroencephalograph (paper speed 25 mm/sec). Signals taken from the output of the driver stages of the Grass were led into electronic integrators which summated the muscle potentials over successive 4-sec. periods, and were recorded as deflections on a second chart drive. The EEG was quantified by passing these signals through band-pass filters (Ross & Davis, 1958) with cutoffs to provide bands of 8–12 cycles/sec and 18–27 cycles/sec. Outputs of the filters were also integrated and recorded on Chart Drive 2 (paper speed: 1 mm/sec).

---

[3] Much of the apparatus used in this experiment was designed and built in this laboratory. Detailed specifications and circuit diagrams of the tracking apparatus, the strain gauge tension induction system, the EEG band-pass filters, and the electronic integrators are on file at the Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, McGill University, Montreal, Canada.

Chart Drive 3 (paper speed: 2.5 mm/sec) was used to record respiration (obtained with a Phipps & Bird pneumograph and recording tambour), palmar conductance, DOT, and the output of the strain gauge amplifier.

Recording electrodes.—All electrodes except those used for palmar conductance were made of cellulose sponge of 1 cubic in. Heart rate was recorded from electrodes placed on the chest wall immediately below the heart, and on the right shoulder. Electrodes for EMG recording were placed on the extensor muscles of the right and left forearms and the calves of the right and left legs (Davis, 1959). Two bipolar EEG leads were taken from the nondominant hemisphere from frontal and occipital positions equivalent to placements C2-F2 and P2-02 of the 10/20 system used at the Montreal Neurological Institute (Jasper, 1941), frontal leads being connected with the EEG lead to the 18–27 cps filters, and occipital leads connected with the 8–12 cps filters.

Palmar conductance (PC) was recorded from the left palm by means of a monopolar method described by Malmo and Davis (1961).

Other apparatus.—The *S*'s chair and tracking apparatus were located in a separate room from that of the amplifying and recording devices. It was a relatively soundproof room, entirely shielded with copper screening to reduce electrical interference, and was supplied with an air conditioning plant for the maintenance of constant temperature. Separating *S*'s room from *E*'s control room was a double wall with a screened observation window. A two-way intercommunication system was employed for instructing *S*s and answering questions.

Procedure

All instructions pertaining directly to the experiment were read to each *S*. When *S* first arrived he was informed that the experiment was designed to study the effects of induced muscular tension on his ability to perform a simple psychomotor task, and that the effects of this tension would be determined by simultaneously measuring certain physiological variables and the muscle tension directly. An assistant then prepared *S* by placing the electrodes as discussed above.

The *S* was then instructed by *E* in the method of inducing tension and in the method of maintaining a given level of tension. Subjective levels of tension were then determined by having *S* squeeze the dynamometer very lightly for 30 sec. After a rest period, this procedure was repeated twice more. The average reading of his pull in kilograms at the end of the three, 30-sec. periods was designated *S*'s Very Light (VL) tension. After another rest period, *S* was asked to squeeze the dynamometer as tightly as he could and to hold this for 30 sec. The reading at the end of the 30 sec. was designated his Maximum (Max.) pull. Following another rest period, *S* was required to subjectively bisect the apparent pull on the dynamometer of the previous two instances (Very Light and Maximum), and to hold this for 30 sec. This also was repeated twice more, and the average of the three dynamometer pulls in kilograms at the end of the 30-sec. periods was designated as *S*'s Very Heavy (VH) tension. Similarly, a Medium (M) tension was determined by bisecting the apparent pull between the Very Light and the Very Heavy, as determined by the average of three readings. Light (L) and Heavy (H) tension conditions of pull were then arbitrarily chosen as the physical midpoints, in kilograms, between the Very Light and Medium, and the Medium and Very Heavy, respectively.

A headset with independent earphones for tracking was placed on *S*'s head and adjusted for comfort. The *S* was then instructed by *E* in the tracking procedure as described above. Each trial consisted of two complete cycles of the tracking pedal being depressed and released. This required 96 sec.; 96 sec. rest was also given between each trial. One practice trial was followed by the learning trials to criterion. The criterion was the tracking of 3 successive trials without improvement, but never less than 6 nor more than 10.

A 10-min. rest period and a glass of water were given between learning and tension trials. No smoking was allowed. During the rest period all equipment was again checked to conform with calibration standards.

The five tension conditions of (VL, L, M, H, and VH) were given once only to each *S* in an order determined by a table of random numbers. The only difference between the tension trials and the learning trials was that during the entire trial *S* was required to maintain the particular tension on the dynamometer in addition to tracking. After two practice trials, *S*s had no difficulty in simultaneously tracking and maintaining a steady level of dynamometer pull.

Two control conditions of No Tension and Exertion (EX) were also randomly presented with the tension conditions. The No Tension control did not differ from the learning trials.

Its purpose was to gauge the effects of performance with no tension. The Exertion condition was introduced in an attempt to determine the effect of induced tension alone (without tracking) on the various physiological measures. This condition required each of 8 *S*s to squeeze the dynamometer at a pressure equivalent to his Very Heavy pull, and each of 25 *S*s to squeeze the dynamometer at a pressure equivalent to his Maximum pull. Five *S*s did not have an Exertion condition at all. Besides squeezing the dynamometer, in the Exertion condition each *S* was required to push the foot pedal down and keep it there for the length of the trial. He also heard a tone in both ears equal in intensity to what he would have heard had he been off target two degrees. Physical exertion was thus at a high level. The primary difference between this and other trials was that there was no actual tracking performance.

To increase the rate of learning during the early part of the experiment, *S*s were told their score in terms of time off target at the end of each trial. This was not continued during the tension trials, however, in order to prevent associations of a particular score with a given level of tension and consequent excitement or activation if the score were relatively poor.

The mean tolerance allowed for the maintenance of tension was about 1.5 kg. In the pilot experiments and the main study, it was found that after a little practice this tolerance was more than sufficient to enable *S*s to maintain tension at quite steady levels over an entire trial.

Treatment of Data

EEG.—Outputs from the band-pass filters were integrated with 4-sec. discharge rates and recorded as deflections (Malmo & Surwillo, 1960, p. 7). During the 96-sec. trial, 24 of these deflections were recorded, and the level of EEG activity for a given trial was then defined as the mean voltage of the 24 integrated deflections.

EKG.—EKG was obtained by counting the beats in each successive 12-sec. period and multiplying by 5 to get the mean number of beats per minute for that 12-sec. period. The eight values thus calculated were averaged for the mean number of heart beats per minute for the trial.

EMG.—EMG being less variable than EEG it was satisfactory to sample only eight peaks during a trial. The measured peaks for the right leg coincided with the deflection occurring approximately midway up and down the rises and falls in muscle tension due to tracking (Deflections 3, 4, 9, 10, 15, 16, 21, and 22 were measured). Deflections were measured with special scales calibrated in microvolts, and the level of muscle activity for a trial was then defined as the mean voltage of the eight measured deflections.

Respiration and palmar conductance.—Respiration rate was found by counting the number of complete inspiration-expiration cycles from one 12-sec. line to the next, correct to the nearest ¼ cycle. The mean rate for the trial was obtained by multiplying this value by 5 to get the number of cycles per minute for that period, and then obtaining the mean of the eight values. Palmar resistance was measured at 12-sec. intervals on the tracing with a special scale calibrated in ohms. This resistance value was converted to its reciprocal, conductance, and the average conductance value in micromhos for the trial was obtained by calculating the mean of the eight conductance measures.

DOT and strain gauge.—The integrated deflections of Distance Off Target and the pull on the strain gauge dynamometer (SG) were measured in the same way as the EEG. That is, each of the 24 deflections was meassured using a special plastic scale, calibrated in degrees for DOT and kilograms for SG. The average of the 24 deflections was taken as the mean DOT or grip pressure for the trial.

Errors.—The number of "up" and "down" errors per trial was obtained directly by means of two electromechanical counters.

RESULTS

Figure 1A shows the continuous and approximately linear rise in force, as measured by the strain gauge dynamometer, through the graded series of induced tensions. Parts B and C of Fig. 1 show that the effect of induced tension in this experiment was to impair tracking performance, with the higher tensions having a more detrimental effect than the lower ones. A two-way analysis of variance of these data shows that the between-trials variance of errors as measured by Distance Off Target was significant ($P = .05$) for the induced tension conditions.

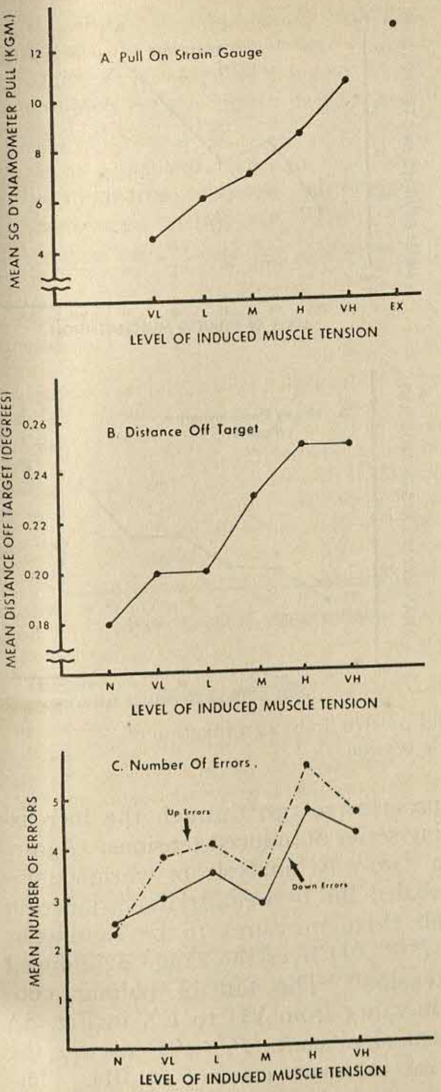


FIG. 1. Performance measures as a function of induced muscle tension.

Figure 2 presents the EMG data. As in the case of the strain gauge, the curve for muscle tension in the right arm, as a function of the amount of tension induced, is approximately linear. The other three curves, for muscles not directly involved in pulling on the dynamometer, likewise show incremental changes corresponding to increasing values of induced tension. In Fig. 2B the right leg is designated as active because it was used in tracking, but the load on the leg muscle was constant, of course, from one induced tension condition to the next. The two left limbs (Fig. 2C and 2D) were not engaged in any activity. Despite this passivity, however, their curves show steady rise through the ascending series of induced tension values up to Cond. VH (i.e., through the series of tensions that were induced during tracking). Analysis of variance for all four muscles also showed the between-trials variance to be significant ($P < .01$) over the range of tension conditions with tracking.

The drop in the left arm in tension from the Cond. VH during tracking to Cond. EX (no tracking) should be especially noted in Fig. 2C. This is shown as a point on the far right. The drop was significant ($P < .01$). It will be recalled that 33 of the 38 *S*s were included in Cond. EX. Twenty-five of these *S*s pulled on the dynamometer their maximum amount during this trial, while 8 *S*s gripped with the same pressure as they used in Cond. VH. With this amount of exertion, the right leg EMG (Fig. 2B) showed a significant rise from VH to EX (the right leg held the tracking pedal all the way down during this condition, exerting the maximum leg tension for a trial). Yet, the right arm EMG (Fig. 2A) which was squeezing the dynamometer, shows only a slight and very nonsignificant rise. In fact, 48% of the 33 *S*s actually showed a fall in the right arm EMG from the VH tracking condition to the EX nontracking condition; 75% of the 33 *S*s showed a fall in the left arm EMG from VH to EX. The slight rise shown in Fig. 2D is not significant,

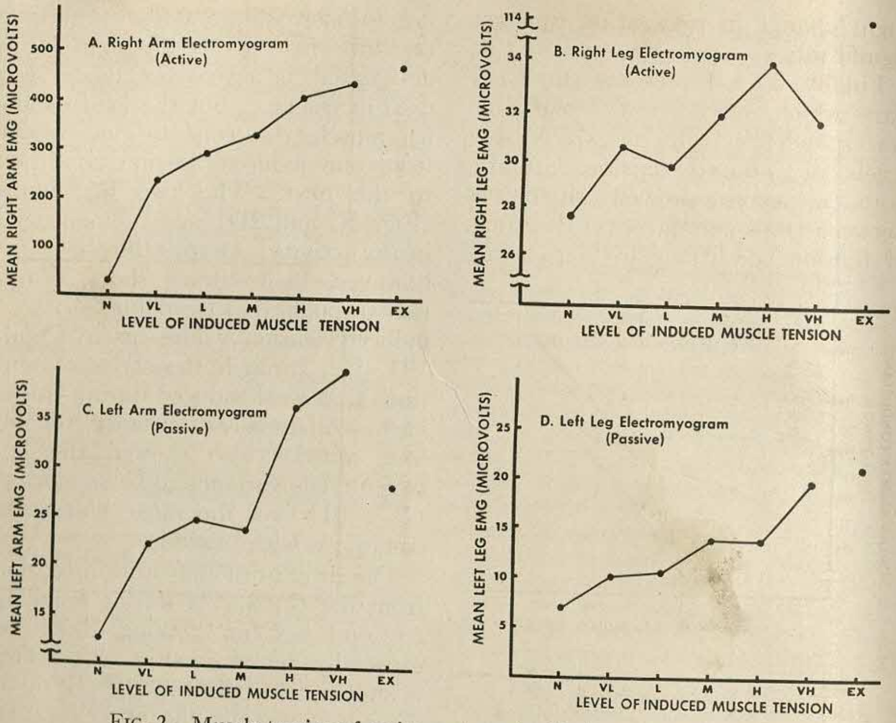


FIG. 2. Muscle tension of active and passive limbs as a function of induced muscle tension.

of course. A summary of the percentage of *S*s whose measures of activation fell for Cond. EX as compared with Cond. VH is shown in Table 1.

Similar curves for palmar conductance, heart rate, and respiration are presented in Fig. 3. It is noteworthy that there are no reversals in the curves which all show a regular progressive rise through the increasing series of induced tensions. Again, a two-way analysis of variance revealed the between-trials variance of all three measures to be significant ($P < .01$) over the range of induced tension. The fall in palmar conductance from VH to EX in Fig. 3A was significant ($P < .05$) as was the rise in heart rate ($P < .01$). The

TABLE 1

PERCENTAGE OF SUBJECTS WITH LOWER VALUES FOR PHYSIOLOGICAL RECORDINGS UNDER THE EXERTION CONDITION THAN UNDER THE TRACKING COND. VH

S's Pull on EX	*N*	Right Arm	Right Leg	Left Arm	Left Leg	Palmar Cond.	Heart Rate	Resp.	EEG 8–12	EEG 18–27
Max.	25	48.00	32.00	72.00	40.00	16.00	0.00	56.00	52.00	60.00
VH[a]	8	50.00	12.50	87.50	62.50	62.50	62.50	75.00	62.50	12.50
Comb.	33	48.48	27.27	75.75	45.45	27.27	15.15	60.60	54.54	48.48

[a] This VH condition is not to be confused with the VH condition during tracking. These 8 *S*s took their Exertion (i.e., no tracking) condition by pulling the same kilogram values on the dynamometer as they had pulled in the VH tracking condition.

slight change in respiration was not significant.

Finally, Fig. 4 presents the EEG data which again show remarkable correspondence with the increasing levels of induced tension. Again, both measures showed significant between-trials variance over the range of tension conditions by a two-way analysis of variance (8–12 cps, $P = .05$; 18–27 cps, $P = .01$). The large reversal between Cond. VH and EX is especially to be noted (i.e., falling 8–12 cps amplitude, significant with $P < .01$, compared with rising 18–27 cps amplitude almost significant at the .05 level).

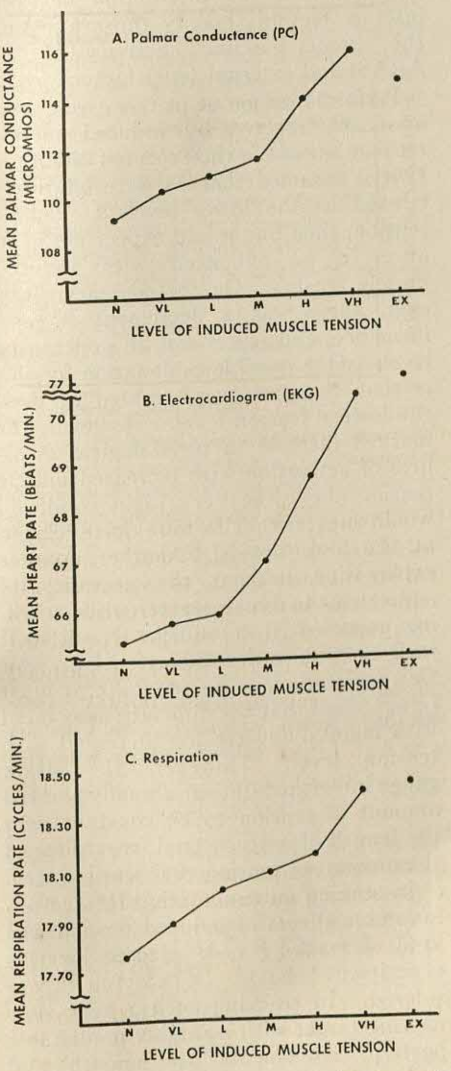


FIG. 3. Peripheral measures of autonomic activity as a function of induced muscle tension.

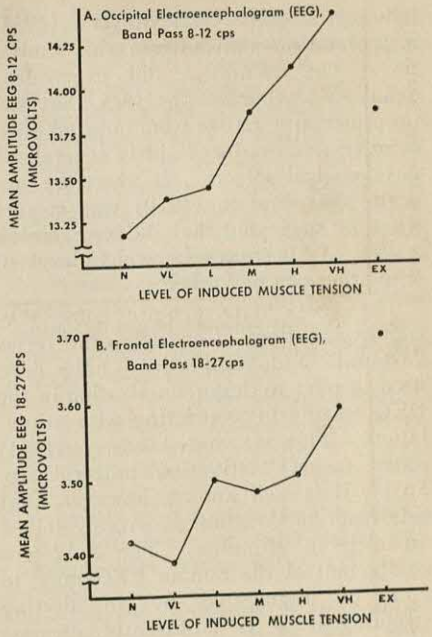


FIG. 4. EEG measures as a function of induced muscle tension.

DISCUSSION

The physiological data were highly consistent in showing regular and continuous rise in level as a function of increments in tension induced in the right arm during tracking. Freeman's (1938) and Freeman and Simpson's (1938) results with palmar conductance were thus confirmed, but the additional data from muscle potentials, respiration, heart rate, and especially EEG considerably extend the possibilities of interpretation. The demonstration of such widespread changes, however, would appear to cast strong doubt on any rationale limited to the skeletal-motor (or somatic) system. It appears that Meyer's (1953) interpretation suffers from this limitation. The results of this experiment definitely encourage the idea that the proprioceptive return from induced muscular tension produces widely generalized physiological effects. It therefore appears that Malmo (1959) was correct when he suggested that induced tension is one of the ways in which level of activation can be varied.

The EEG data require some special explanation. Generally, the terms "arousal" and "activation" have been used to refer to desynchronization in the EEG tracing in association with stimulation. This is what Lindsley (1951) refers to as "activation pattern" (p. 505). It is well known, however, that this reaction does not inevitably follow an alerting stimulus. The 8–12 cps component of the human EEG may, in fact, be augmented by an alerting stimulus if *S* is sufficiently drowsy. Stennett (1957) has systematically investigated the direction of the 8–12 cps change as a function of the *S*'s prior level of activation showing it to be an inverted U function.

In the present experiment, Stennett's (1957) practice of obtaining average voltage by means of band-pass filters was employed. This method permitted independent measurements of 8–12 cps EEG amplitude and 18–27 cps EEG amplitude. While the overall change in 8–12 cps EEG might have been expected to occur in the downward direction because of the generally activating effects of induced tension, finding a change in the upward direction was by no means unprecedented. In addition to Stennett's results there are recent (unpublished) findings of Malmo[4] showing a significant overall rise in 8–12 cps amplitude in going from a condition of lower, to one of higher, activation level. Malmo and Surwillo (1960) have also shown that under appropriate conditions an overall fall of 8–12 cps EEG can occur with a rise in activation level.

The evidence from Cond. EX (Table 1) quite clearly suggests an interaction between tension induction and tracking. That is, if induced tension were the sole factor, the points on the curves for Cond. EX should be more continuous with the points for Cond. VH than they in fact are. These results support the previous contention of Malmo (1959, p. 373) that activation level is invariably determined by an interaction between internal factors (mainly those affecting the "tonic baseline" activity of the ARAS) and external (cue) factors.

Performance per se in this experiment was also affected by induced muscle tension but not in the expected direction. If it is assumed that *S*s were relatively relaxed at the lower levels of muscle tension, then one would expect performance to be enhanced with induced muscle tension. On the contrary, however, there was a decrement in performance, and scores rose to prelearning levels. One possible explanation for this is that *S*s were overactivated at even the lowest tension level. However, the marked increase in physiological measures of activation with increased muscle tension above the Very Light condition would suggest that *S*s were quite relaxed at the lower levels. Another possible explanation is that the necessity of squeezing the dynamometer while tracking represented a sufficiently difficult task to require divided attention. For this explanation to be adequate, it must be supposed that the difficulty associated with maintaining tension increased with tension level. However, the strain gauge integrated output actually showed amount of tension to be constant over the length of a given trial, regardless of the amount of tension that was induced.

In seeking an explanation for the unfavorable effects of induced tension, the kind of tracking task selected for this experiment should be carefully considered. In tracking such a very slow moving target with its highly predictable pattern, movements are smooth and relaxed in contrast to the high tension movements required to do well in tracking a high frequency, irregular target course under visual control. It seems quite possible that had the latter (high tension) kind of task been selected the results may have been different, with induced tension improving performance rather than degrading it.

---

[4] R. B. Malmo, personal communication.

SUMMARY

Induced muscle tension has been demonstrated to have significant effects in relation to a wide range of behavioral phenomena. From the behavioral evidence, it appeared reasonable to consider that induced tension was only one of the many ways in which activation level can be varied. It followed that if induced tension were a reliable means of varying activation level, then in addition to the behavioral effects which have already been demonstrated, induced tension should also produce regular and consistent changes in the various physiological indicants of activation. The purpose of this experiment was to investigate such physiological changes.

Thirty-eight male college students were trained in an auditory tracking task. Physiological indicants of activation included heart rate, respiration rate, palmar conductance, frontal and occipital EEG, and EMGs from active and passive limbs. Following learning trials, *S*s were required to track while squeezing a hand dynamometer in order to maintain one of five predetermined levels of muscle tension.

Results clearly showed close agreement between amount of tension induced and the level of activity in all physiological measures. These results were considered in support of a theory that the proprioceptive return from the induced muscle tension produces generalized behavioral and physiological effects indirectly by increasing activity in the reticular activating system.

REFERENCES

COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347–367.

DAVIS, J. F. Manual of surface electromyography. *USAF WADC tech. Rep.*, 1959, No. 59–184.

DAVIS, J. F., STENNETT, R. G., & QUILTER, R. E. An auditory tracking device designed for use in conjunction with continuous EEG recording. *Percept. mot. Skills*, 1957, **7**, 239–244.

DELAFRESNAYE, J. F., ADRIAN, E. D., BREMER, F., & JASPER, H. H. (Eds.), *Brain mechanisms and consciousness.* Springfield, Ill.: Charles C Thomas, 1954.

FREEMAN, G. L. The optimal muscular tensions for various performances. *Amer. J. Psychol.*, 1938, **51**, 146–150.

FREEMAN, G. L., & SIMPSON, R. M. The effect of experimentally induced tension upon palmar skin resistance. *J. gen. Psychol.*, 1938, **18**, 319–326.

JASPER, H. H. Electroencephalography. In W. Penfield & T. C. Erickson (Eds.), *Epilepsy and cerebral localization.* Springfield, Ill.: Charles C Thomas, 1941. Pp. 380–454.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

MALMO, R. B., & DAVIS, J. F. A monopolar method for the measurement of palmar conductance. *Amer. J. Psychol.*, 1961, **74**, 106–113.

MALMO, R. B., & SURWILLO, W. W. Sleep deprivation: Changes in performance and physiological indicants of activation. *Psychol. Monogr.*, 1960, **74**(15, Whole No. 502).

MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.

ROSS, W. R., & DAVIS, J. F. Stable bandpass filters for electroencephalography. *IRE Canad. Conv. Rec.*, 1958, No. 860, 202–206.

SCHNORE, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *J. exp. Psychol.*, 1959, **58**, 117–128.

STENNETT, R. G. The relationship of alpha amplitude to the level of palmar conductance. *EEG clin. Neurophysiol.*, 1957, **9**, 131–138.

(Received August 27, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 532–533

REPLICATION REPORT: TWO FAILURES TO REPRODUCE EFFECTS OF ANXIETY ON EYELID CONDITIONING

MARGARET S. KING, GREGORY A. KIMBLE
Duke University
JOHN GORMAN, AND RICHARD A. KING
University of North Carolina

Taylor (1951), Spence and Taylor (1951), Spence and Weyant (1960), and Baron and Connor (1960), among others, have shown that highly anxious *S*s develop conditioned eyeblinks at a faster rate and to a higher asymptote than less anxious *S*s. The two experiments reported here are practically identical in methodology with those mentioned above. They failed to show that highly anxious *S*s condition better than less anxious *S*s.

Experiment I[1]

Method.—This experiment investigated the relation between anxiety and the presence or absence of a ready signal in eyelid conditioning. The *S*s were chosen from introductory psychology classes on the basis of extreme scores on the Manifest Anxiety (*MA*) scale. Those who obtained scores from 24–42 were designated as high-anxiety *S*s; low-anxiety *S*s were those who obtained scores from 1–7. One-half of the *S*s in each of the anxiety groups received a verbal ready signal 2, 3, or 4 sec. prior to the CS light; the other *S*s did not receive the ready signal. There were 16 *S*s, 8 male and 8 female, in each of the four experimental groups. The apparatus, method of scoring CRs, and procedure were similar to that described by Dufort and Kimble (1958) and Spence and Taylor (1951).

Results.—The results are shown in Fig. 1. As can be seen, there is (*a*) no difference between the performance of the high-A and low-A groups which did not receive a ready signal; (*b*) a tendency for low-A *S*s who received a ready signal to perform better than high-A *S*s who received a ready signal (difference not significant on Trials 71–80); and (*c*) both high-A and low-A groups which received a ready signal performed at significantly lower levels than the no-ready groups ($F = 22.39$, $P < .001$, $df = 1/56$). This was determined by an analysis of variance in which ready signal, anxiety, and sex were the factors.

The first finding, where there was no ready signal, is contrary to those of Spence and Weyant (1960) and Baron and Connor (1960). The second finding, where there was a ready signal, is contrary to those of Taylor (1951) and Spence and Taylor (1951). The third finding is in accord with one made by Dufort and Kimble (1958).

Experiment II[2]

Method.—This experiment investigated the relation between the *MA* scale measure of anxiety, the Welsh (1956) *A* measure of anxiety, and eyelid conditioning. The *S*s were chosen from introductory psychology classes on the basis of high and low *MA* scale and Welsh *A* scores. On the *MA* scale, *S*s having scores of 24–36 were high-anxiety *S*s, while those having scores below 7 were low-anxiety *S*s. On the Welsh *A* scale, *S*s with scores above 22 were put in the high-anxiety group, while *S*s with scores below 6 were put in the low-anxiety group. There were 15 male *S*s in each of the *MA* scale groups (high *MA*

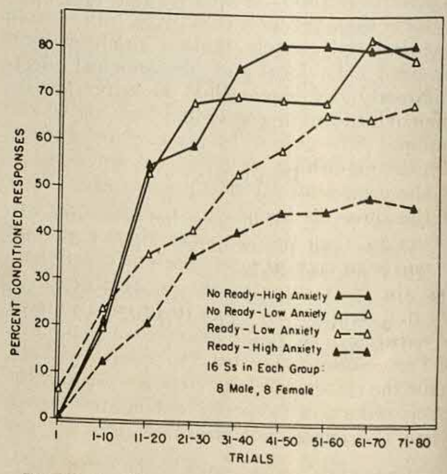


FIG. 1. Conditioning curves for the four groups of Exp. I.

---

[1] This experiment was performed by Margaret A. King in partial fulfillment of the requirements for the MA degree at Duke University under the direction of G. A. Kimble.

[2] This experiment was performed by John Gorman in partial fulfillment of the requirements for the MA degree at the University of North Carolina under the direction of R. A. King.

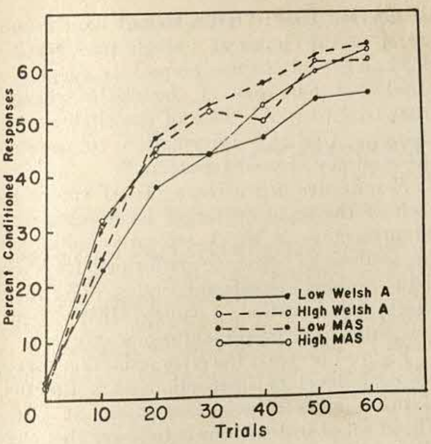


FIG. 2. Conditioning curves for the four groups of Exp. II.

scale, low *MA* scale) and 20 male *S*s in each of the Welsh *A* groups (high Welsh *A*, low Welsh *A*). The apparatus, method of scoring CRs, and the procedure were similar to that described by Dufort and Kimble (1958) and Spence and Taylor (1951).

Results.—The results are shown in Fig. 2. As can be seen, there is (*a*) no significant difference between the high and low *MA* scale anxiety groups; and (*b*) no significant difference between the high and low Welsh *A* anxiety groups.

The contradictory findings presented here make it likely that anxiety as measured by the *MA* scale or the Welsh *A* scale is an interacting variable and has an effect upon conditioning only when it is combined with the proper values of other variables not explicitly manipulated by *E*.

REFERENCES

BARON, M. R., & CONNOR, J. P. Eyelid conditioned responses with various levels of anxiety. *J. exp. Psychol.*, 1960, **60**, 310–313.

DUFORT, R. H., & KIMBLE, G. A. Ready signals and the effect of interpolated UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **56**, 1–7.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SPENCE, K. W., & WEYANT, R. G. Conditioning performance of high and low anxious *S*s in the absence of a warning signal. *J. exp. Psychol.*, 1960, **60**, 146–149.

TAYLOR, J. A. The relationship of anxiety to the conditioned response. *J. exp. Psychol.*, 1951, **42**, 81–92.

WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

(Received February 25, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 533–534

SUPPLEMENTARY REPORT: EFFECTS OF INSTRUCTIONS ON THE PERCEPTION OF MULTIPLE TARGETS[1]

HARRY W. KARN AND LEE W. GREGG

Carnegie Institute of Technology

Karn and Gregg (1961) have previously reported a form of perceptual learning which required *S* to detect the presence or absence of a distinguishing binary cue, a dot, in each of three circles in a triangular array. Two of the three circles provided unchanging cue values in that the dot was always present in one location and never present in another. For the third circle, presence or absence of the dot was random. As a function of amount of experience with the complex at exposures of 1 sec.—ample time for *S* to observe without error the state of the three circles—significant error reduction occurred during test series when the exposure time of .1 sec. was too short to permit scanning of the entire stimulus field.

It was anticipated that *S*s would rationally attack the perceptual task by learning the nature of the unchanging elements of the complex and then direct attention toward the changing element. Nearly perfect performance should ensue. Instead, the learning which occurred was slow and incremental and appeared to be the result of some kind of set. If cognitive sets play a critical role, performance should be dependent upon instructions given *S* concerning the task requirements. A simple suggestion or hint can result in insightful solutions in certain instances (Luchins, 1942; Maier, 1933).

As an initial attempt to provide a better understanding of the behavior generated by

---

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-770.

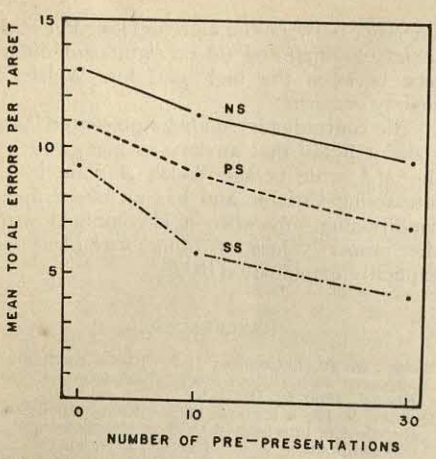


FIG. 1. Mean total errors per target as a function of number of prepresentations at 1 sec. for nonspecific (NS), problem solving set (PS), and specific solution groups (SS).

the multiple stimulus task the instructions were modified to induce a problem solving set at two levels of explicitness.

Method.—The apparatus, general procedure, and group allocation of *S*s were the same as those of the previous study (Karn & Gregg, 1961). Three configurations (Cs) defining the specific combinations of loading (presence, absence, or random occurrence of a dot in a circle) and location in the triangular array were used. These Cs form a latin square so that each loading occurs once and only once at each location. Amount of prior experience was varied by presenting 30, 10, or 0 prepresentations of the particular C at 1 sec. exposure. There followed 60 presentations of the same C at .1 sec.

The first set of instructions were nonspecific (NS) in that *S* was informed that sometimes one or more of the circles would contain a dot. A report on all circles was required.

To induce a problem solving set (PS) the following was added to the basic instructions. "Even though you cannot see all of the circles at a single presentation, it is possible to get all of them correct. Try to discover how this can be done."

The third set of instructions were worded in such a way that a specific solution could be inferred (SS): "Even though you cannot see all of the circles at a single presentation, you can get all of them correct by keeping in mind that only one of the circles changes from trial to trial. Two of the circles never change. Use this information to improve the accuracy of your reporting."

Results and discussion.—Total errors for each of the separate target locations during 60 presentations at .1 sec. were obtained. An analysis of variance performed for these data yielded significant ($P < .01$) main effects for Location, Loading, Instructions, and Number of prepresentations at 1 sec.

Figure 1 presents the error reduction curves for each level of instruction as a function of the number of prepresentations at 1 sec. The effect of instructions is to lower the overall magnitude of errors. Differences in the slopes of these curves are not significant; rate of error reduction remains essentially constant.

The failure to obtain insightful shifts in performance suggests that the perceptual processing which occurs remains at the level of visual discrimination based on criteria as yet unspecified. The lack of evidence for a rational processing of the information inherent in the stimulus complex as a function of problem solving sets, does not mean that learning sets should be discounted or minimized in explaining the behavior. A variety of unique learning sets have been suggested or implied as playing critical roles in perceptual learning situations. Nonverbalized perceptual and motor sets are possibilities as are related eye movement biases in search tasks. The precise ways in which these factors govern the behavior under investigation remains to be determined. Whatever their roles, they appear to have a hampering influence on the problem solving process envisaged in terms of a selective utilization of information provided by the stimulus complex.

REFERENCES

KARN, H. W., & GREGG, L. W. Acquisition of perceptual responses as a function of loading, location, and repetition. *J. exp. Psychol.*, 1961, 62, 62–69.

LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, 54(6, Whole no. 248).

MAIER, N. R. F. An aspect of human reasoning. *Brit. J. Psychol.*, 1933, 24, 144–155.

(Received March 13, 1961)

Journal of
Experimental Psychology

Vol. 62, No. 6 December 1961

CONTRAST EFFECTS IN THE VERBAL CONDITIONING OF MEANING[1]

FRANCIS J. DI VESTA

Syracuse University

The present study extends the findings of earlier investigations of contrast and assimilation effects to attitude change accomplished through semantic mediation processes. The rationale is based principally on the concept-formation approach to attitude acquisition (Rhine, 1958). A basic assumption in this orientation is that a concept is defined by some principle or common dimension that permits the classification of objects. The principle is abstracted as a feature of a class of stimulus patterns through recurrent experiences. Such classifications can be made on the basis of both the objective and evaluative characteristics of the dimension of meaning. Thus, in addition to whatever objective characteristics contribute to the formation of a concept, it is conceivable that pleasant experiences might add the dimension of *good* to the meaning of a concept while unpleasant experiences or adverse information from a trustworthy source might add the dimension of *bad* to the concept. For all objects and experiences an essential condition for concept formation is the associations between a common response and a variety of stimuli.

Among the complex behavioral processes, the naming response is a most important basis for a common response. A simple example is that a common name can be attached to a series of objects, all of which have different physical characteristics to form a *class*, for example, the word *furniture* may be applied to many different objects which are physically quite dissimilar. The name, or verbal response, provides the basis for a common mediating process. As mediators they produce stimuli which may then become conditioned to other mediating, instrumental, or attitudinal responses, thereby permitting the facilitation of new learning whether cognitive or affective. Demonstration of the feasibility for producing attitudinal change within this theoretical framework has been provided through the use of what Staats and Staats (1957, 1958) refer to as "classical

---

[1] This research was supported by Research Grant M-2900 from the National Institute of Mental Health, United States Public Health Service. Thanks are due to Marilyn Rothschild and Fred W. Naugle for their invaluable assistance in collecting the present data.

conditioning." The effects of differential reinforcement on attitude strength, development, and change has been demonstrated by Rhine and Silun (1958). That evaluative stimulus components comprise one dimension of meaning which mediate verbal responses in forming associations is documented by the factorial analyses of meaning (Osgood, Suci, & Tannenbaum, 1957).

The research designs used in the studies exemplified by those of Staats and Staats (1957), as well as that of Rhine and Silun (1958), typically followed the pattern of developing two attitudes concurrently. For example, one concept-attitude was developed by pairing one name with words having positive-evaluative meanings and a second concept-attitude was developed by pairing another name with words having negative-evaluative meanings; or, one of the aforementioned was used in combination with the development of a "neutral" attitude. Treatment effects were determined by comparing the intra-individual changes or difference between the two attitudes following the experimental treatments.

In none of these studies was the possibility of a direct-opposite effect considered as an alternative for the formation of one of the concepts. In view of much of the recent (as well as the classical) literature on contrast effects we would hypothesize that similar effects will occur in the development of attitudes by conditioning. Changes are to be anticipated under these circumstances, in attitudes concerning related issues as well as the specific target issue toward which the persuasive effort is directed. In similar fashion, the literature in which a contrast effect has been demonstrated in judgments of simple stimulus materials (Harvey & Caldwell, 1959; Sherif, Taub, & Hovland, 1958) suggests the occurrence of a similar phenomenon in the simultaneous development of two attitudes. Under these circumstances, the more intense of the two stimuli serves as an anchor or reference point against which the other stimulus (attitude object) is evaluated. While the changes may either accompany the change in the target attitude or may occur instead of it, it is the former condition that is of present concern.

Thus, if contrast effects are to be demonstrated in the evaluation of a poorly structured concept, polarization will occur toward a position opposite to that evaluation made of the anchoring (target) concept-attitude. The degree of polarization thus will be inversely related to the scale position of the anchor object. The more remote the anchors the greater will be the displacement of the scale value of the related attitude from that of the anchor.

The general hypothesis is that a contrast effect is obtained when two attitudes are formed concurrently through verbal conditioning. The more specific hypothesis is that if the target concept-attitude is formed on the basis of a conceptual dimension such as evaluation, potency, or activity, the change in the *accompanying* (neutral) concept will also be on that dimension but toward the opposite pole. Since the concept dimensions of potency and activity also have high factorial loadings on the evaluative dimensions, it is assumed that a change in one of these concept dimensions will also be accompanied by a change in the evaluative dimension.

These general assumptions are suggested in experiments where judgments have been made on simple stimulus materials (Sherif et al., 1958), and in experiments where both source and content were studied jointly (Hovland, Harvey, & Sherif, 1957) as variables influencing reactions to persuasion. None of the aforementioned studies has been concerned with contrast effects in laboratory situations where complex verbal stimuli were used in the development of attitudes, or where the development of one attitude was modified by the development of the other, while simultaneously minimizing or eliminating the complicating effects of source.

METHOD

Subjects.—Seventy-four students participated in the experiment as *S*s to fulfill one of the course requirements in introductory psychology. There were 16 *S*s in each of four experimental groups and 10 *S*s in the control group.

Procedure.—The general procedure, with but slight modifications, was that employed by Rhine and Silun (1958) and also described by Rhine (1958). Briefly, it was as follows: Subjects were administered the task individually by use of a set of 50 4 × 6 in. cards on each of which was a single word, in primary type, representing a "trait" or "characteristic." The instructions were that *E* wanted to find out whether *S*s could make predictions about the characteristics of each of two little known cultures from the kinds of information typed on the cards. Since *E* wanted *S* to start out with as little information as possible the actual names would not be used, but instead the two words CEB and LOJ would be used to label the cultures. The "names" were picked arbitrarily from Glaze (1928) association values of 27%. The *S* was told to look at one trait at a time and predict to which culture it belonged by first saying the name of the culture and then the trait (e.g., *S* might say, "CEB-GRANDIOSE"). Following each such prediction *E* announced the correct combination as a basis for *S* to evaluate his performance. If *S* was correct he went on to the next card. If incorrect, *S* repeated the correct combination and then proceeded with the next prediction. The *S* went through 50 trait names in order and then repeated the series a second time. A record of *S*'s errors was kept by *E*.

Following this procedure *S* rated each of the newly developed concepts, CEB and LOJ, separately on 20 semantic differential scales in exactly the same form and order used by Jenkins, Russell, and Suci (1958, p. 690). These scales had been selected by Jenkins et al. (1958) because the factorial loadings were highest on the dimension sampled by each scale and low on all other factors.

The treatments differed with respect to the concept-formation tasks required of the experimental groups. The general format of the task, however, was exactly the same for all groups. It was comprised of 50 different trait names from which *S* was to make his prediction; 25 of these traits were to be associated with one of the nonsense syllables and were classifiable on either the evaluation or potency dimension depending upon the experimental variation in which the list was to be used; the remaining 25 traits were represented by neutral words and were not classifiable on the respective dimensions of meaning. The specific words and the serial position of each neutral word were exactly the same for each of the experimental groups and the control group. The position occupied by the traits classifiable on the evaluative or potency dimension was the same in all groups. For the control group, another set of neutral words was used in place of the experimental variations of the word lists.

The nonsense syllable used to designate the critical concept was associated in Group C_{NE} with unfavorable (negative-evaluative) words; in Group C_{PE} with favorable (positive-evaluative) words; in Group C_{HP} with words high in value on the potency dimension; in Group C_{LP} with words low in value on the potency dimension; and, in the control group (Group C_N) with words having neutral meanings. In all groups, as indicated above, the other nonsense syllable was associated with words having neutral meanings, the same words being used for all groups. The data related to the learning of the neutral words will be designated by the symbols N_{NE}, N_{PE}, N_{HP}, N_{LP}, and N_N for the respective experimental and control groups.

In the positive evaluative list were included such words as GENUINE, RELIABLE, and CHARITABLE and in the negative evaluative list were such words as DESTRUCTIVE, DIRTY, and CHEAT. The words in both of these lists were selected from those used by Rhine[2] and Silun who chose only those words for which there was 90% or more agreement among *S*s on the classification of the words as "very desirable" or "very undesirable" traits.

---

[2] R. J. Rhine, personal communication.

The words used in the high and low potency meaning lists and in the neutral list were selected on the basis of mean concept values on the masculine-feminine and hard-soft scales published by Jenkins et al. (1958). Words for the high potency list were selected if the mean values approximated 3.00 or less. The average scale values of all words finally used were 2.24 and 1.96, respectively, on the two scales. The opposite ends of the same scales were used in the selection of words with low potency meaning. The average scale values for the 25 low potency words on the two scales were 5.19 and 5.36, respectively. Illustrations of words having high potency meanings are RAPID, BASE, ANGER, and FIRE; of words having low potency meanings are SMOOTH, LEISURELY, and CALM.

Examples of words in the neutral list are FAR, ARCHED, LINGERING, and CONTINUOUS. The words in this category were chosen from the range of *Atlas* values between approximately 3.00 and 5.00 on each of the three dimensions concerned, and insofar as possible on the remaining 14 scales. In practice, there were few departures from this general rule. The average scale values of the words so selected were 3.87 on the evaluative scales, 4.05 on the potency scales, and 3.86 on the activity scales.

The concepts selected for each list thus were chosen from a specified portion of the rating scale for the manipulated dimensions. However, as noted by the authors (Jenkins et al., 1958), and as is evident from the examples above, the *Atlas* does not permit the selection of words representing factorially pure meanings.

The analyses to be reported below are based on *S*s' ratings on 6 of the 20 scales used. The cruel-kind and good-bad scales were used in measuring the evaluative dimension, and the hard-soft and masculine-feminine scales for measuring the potency dimension. These particular scales were selected because they represent the most highly loaded dimensions involved in learning the critical concepts by the respective experimental groups. The active-passive and excitable-calm scales were used, in addition, to measure the oriented-activity dimension. Inasmuch as no attempt had been made to vary the activity dimension it was assumed that comparisons of data from the evaluative and potency scales with that gathered from this measure might provide a supplementary control. In practice, however, it was virtually impossible to select words from the available lists which were high on either the potency or evaluation dimension and low on the activity dimension. In all comparisons, analyses, and combinations of scales within a dimension, the favorable pole of the scales was aligned with a rating of 7 and the unfavorable pole with a rating of 1.

The nonsense syllable associated with the critical concept and that with the neutral concept were counterbalanced within each group. In none of the groups were differences in learning, and in the consequent ratings for concepts learned to the syllables CEB and LOJ, significant ($P > .05$) by *t* test. Since the nonsense syllables were found not to have a significant effect on either learning or ratings the two subgroups were combined

TABLE 1

MEAN RATINGS COMPARED WITH MEAN ATLAS VALUES OF WORDS USED IN CONDITIONING THREE CONCEPT-ATTITUDES

Semantic Differential Scale	Concept-Attitude					
	Neutral (N_N)		High-Potency (C_{HP})		Low-Potency (C_{LP})	
	Mean Rating	Mean *Atlas*	Mean Rating	Mean *Atlas*	Mean Rating	Mean *Atlas*
Cruel-Kind	4.4	4.2	2.3	2.9	6.1	5.9
Good-Bad	3.4	3.5	4.2	4.0	2.1	2.3
Masculine-Feminine	3.9	4.0	1.7	2.2	5.9	5.2
Hard-Soft	4.6	4.1	1.9	2.0	5.5	5.4
Active-Passive	2.2	3.9	1.7	2.2	3.5	3.9
Calm-Excitable	4.0	3.8	5.5	5.0	1.9	2.8
D	1.80		1.08		1.24	

Note.—In this and the following tables, *N* in Group N_N is 10, and *N* in each other group is 16.

TABLE 2

Means and *SD*s of Concept Ratings on Evaluative, Potency, and Activity Scales

Concept-Attitude	Rating Scale					
	Evaluative		Potency		Activity	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Neg. Evaluative (C_{NE})	1.44	0.58	1.72	0.72	2.07	1.12
Neutral (N_{PE})	3.34	1.22	3.15	1.37	3.69	1.63
Neutral (N_N)	4.80	1.55	4.25	1.46	3.05	1.21
Neutral (N_{NE})	5.88	0.80	5.19	1.02	5.59	1.27
Pos. Evaluative (C_{PE})	6.56	0.71	4.15	1.39	4.09	1.24
High Potency (C_{HP})	3.03	1.56	1.76	1.20	1.99	1.30
Neutral (N_{LP})	3.69	1.05	2.69	1.23	3.31	1.58
Neutral (N_N)	4.80	1.55	4.25	1.46	3.05	1.21
Neutral (N_{HP})	4.86	1.52	5.26	1.10	4.99	1.63
Low Potency (C_{LP})	6.00	0.81	5.72	1.14	4.79	1.25

within each experimental group and in the control group for all subsequent analyses.

Results

In Table 1 *S*s' ratings, on six semantic differential scales, of C_{HP}, C_{LP}, and N_N are compared with the mean of the *Atlas* values for all 25 words that were used in the conditioning of each attitude in these groups. The difference, d, between the mean ratings of the conditioned concept-attitudes and the mean *Atlas* values was found for each of the scales. These were combined into a single measure, D, a descriptive statistic, of the distance between the conditioned concept-attitude and the *Atlas* scale values using the formula $D = \sqrt{\Sigma d^2}$. While the significance of D cannot be tested directly it should be noted that the maximum value of D with the six scales is 14.7. Thus, the small values reflect essential agreement between *S*s' ratings and those published in the *Atlas*. Inasmuch as the negative-evaluative and the positive-evaluative lists were obtained from another source, comparable comparisons were not possible.

In Table 2 are presented the means and *SD*s of ratings given the critical and neutral concepts on each of the three dimensions, evaluative, potency, and activity, by each group. The ratings on the two individual scales, on which each dimension was measured, were combined for this and all subsequent analyses to obtain a total "factor" score. Similarly, since the difference in ratings between C_N and N_N was not significant ($P > .05$, sign test) the two were combined and are represented as N_N.

While there is considerable generalization among factors to be noted in the data summarized in Table 2, the most consistent and strongest effect occurs in the factor or dimension on which the word lists had been selected. Also apparent is the effect on the ratings of the neutral concept by the principle or dimension involved in the conditioning of the critical concept. The scores are arranged to highlight this effect, the arrangements being identical for the two dimensions.

In Table 3 are presented the distances for each critical concept compared with the relevant neutral con-

TABLE 3

Distances (D) between Ratings Made of the Critical and Neutral Concepts

Critical Concept	Distance (of Critical Concept) from Neutral Concept Learned Concomitantly With:		
	Critical Concept	Neutral Concept	Opposing Critical Concept
C_{HP}	7.44	4.36	2.37
C_{LP}	6.12	3.97	2.99
C_{PE}	5.24	3.15	3.07
C_{NE}	8.79	6.27	4.01

cepts (either that learned with the critical concept or with the opposing critical concept). In every instance, the lowest values of D are between the critical concepts and the neutral concepts conditioned with the opposing critical concept. The largest values of D, on the other hand, are between the critical concepts and the respective neutral concepts conditioned with them. The data in this table reflect the specific effects of the meaning dimension on the ratings of the concept. In every instance where the difference was not significant ($P > .05$) it was either on the activity scale or on the scale for the unrelated factor.[3]

The comparisons among the neutral concepts learned in conjunction with the various critical concepts are presented in Table 4. It is clear, from this as in the previous table, that the neutral concept is strongly affected by the concurrent learning of another concept with a strong meaning dimension. The largest Ds are to be found between the neutral concepts learned under opposing conditions.

All Ds shown in Tables 3 and 4 were tested indirectly on each dimension by use of nonparametric statistics.[3] The sign test was used for all related samples (e.g., in testing the significance of the difference between ratings of C_{NE} and N_{NE}). The Fisher exact probability test was used for testing the difference between ratings of all independent samples (e.g., in testing the difference between the ratings of the C_{NE} and N_{PE}). The median for all comparisons was between the .01 and the .001 level of significance. These differences indicate: (*a*) that the critical concepts were developed in accordance with the meaning dimension of the words that were associated with them; (*b*) that, on the dimension being conditioned, the neutral concepts were substantially different from each other; and (*c*) the ratings made of the neutral concepts were most like the ratings of the opposite critical concept.

A further analysis demonstrating the dependence of the rating of the neutral concept on the rating of the respective critical concept was accomplished by correlating the two ratings of all *S*s ($N = 74$) in all groups. The Pearson r's, all high and significantly greater than zero ($P < .01$), are as follows: $r = -.73$

TABLE 4

Distances (D) between Ratings Made of the Neutral Concept Learned under the Several Experimental Conditions

Comparison	D	Comparison	D
N_N vs. N_{HP}	3.75	N_N vs. N_{NE}	4.29
N_N vs. N_{LP}	2.62	N_{LP} vs. N_{HP}	5.17
N_N vs. N_{PE}	3.44	N_{NE} vs. N_{PE}	4.78

---

[3] Two tables presenting the levels of significance on each dimension for the comparisons shown in Tables 3 and 4 have been deposited with the American Documentation Institute. Order Document No. 6863 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

for evaluative ratings; $r = -.79$ for potency ratings; and $r = -.76$ for activity ratings.

Differences between experimental groups in the total number of errors made during the learning phase were tested for significance by analysis of variance for simple randomized groups. The mean number of errors for each group and the F for the comparison among groups are presented in Table 5. The comparisons among means were made by Duncan's multiple range test. These comparisons revealed no significant differences in learning errors between Groups C_{NE} and C_{PE}, or between Groups C_{HP} and C_{LP}, on the critical concepts. Group C_{PE} did learn with significantly ($.05 > P > .01$) fewer errors than either Group C_{HP} or Group C_{LP}, however. In the comparisons involving the learning of the neutral concepts, significantly ($P < .05$) fewer errors were made in learning N_{PE} than in learning N_{HP}, N_{LP}, or N_{NE}. No other differences were found to be significant. Since these differences were not related in any systematic way to the concepts being compared, it is doubtful that the degree of learning affected the ratings made. On the other hand, in every instance, the critical concept was learned with fewer errors than was the neutral concept. Over all groups, approximately 40% more errors were made in learning the neutral concepts, running as high as twice as many errors in the groups learning words varied on the evaluative meaning dimension. In the control group ($N = 10$) an average of 23.6 ($SD = 4.63$) errors were made in learning one of the neutral concepts and an average of 23.7 ($SD = 3.82$) errors in learning the other neutral concept. This result is not significantly different from expectations based on chance.

TABLE 5

MEAN ERRORS IN LEARNING THE CRITICAL AND NEUTRAL CONCEPTS IN EACH EXPERIMENTAL GROUP

Group	Critical Concept		Neutral Concept	
	Mean	*SD*	Mean	*SD*
C_{NE}	8.50	7.39	15.81	6.03
C_{PE}	5.52	4.97	11.44	4.47
C_{HP}	13.44	4.58	17.19	3.42
C_{LP}	11.13	6.17	18.81	6.52
$F(df = 3/60)$	4.73		5.34	
P	$<.01$		$<.01$	

DISCUSSION

The evidence supports the general hypothesis that attitudes can be developed by associating symbols (nonsense syllables) with words having in common an explicit meaning dimension. Here the assumption is that through the past history of *S* positive-meaning words were associated with some mediator like that representing good and negative-meaning words with some other mediator such as bad. Similarly, high potency words are, by a similar process, associated with some common mediator like hard and low potency words with a mediator such as soft. In the present experiment the evaluative dimension of the stimulus-traits was most distinct for the groups learning the negative or positive characteristics and the potency dimension was most distinctive for the remaining two groups. During the learning process it is assumed that the abstraction by *S* of the common element mediated the rating of the critical concepts on the semantic differential scales. Our results thus support those of previous investigators, particularly the conclusions of Rhine and Silun (1958). At this point in the discussion, the major difference between the present study and the earlier investigations is that in the present study the procedure was extended to dimensions of meaning other than the evaluative dimension. Thus, it was demonstrated that the potency dimension of meaning identified by Osgood et al. (1957) does provide a basis for semantic mediation as does the evaluative dimension. In addition, the semantic differential scales were used in rating the concepts, a procedure closer to that traditionally used in the measurement of attitude than the continued elicitation of a successive series of traits used by Rhine and Silun (1958).

The extreme stability of the ratings of the various concepts developed can almost be described as dramatic. As indicated, the present ratings were nearly identical with those arrived at through an average of the *Atlas* values of the words used in conditioning both potency and neutral meaning to the nonsense syllables. Furthermore, in a very real sense, our four experimental groups comprise four replications. In each case, there was no question that the ratings of the critical concepts were in the direction and of the approximate magnitude that would be predicted on the basis of the words used.

The major hypothesis of the present study was that the ratings made as a consequence of the conditioning of the critical concept would be inversely related to the ratings made of the syllable associated with the neutral words. This hypothesis was supported completely in each of the experimental groups. Not only is such a contrast effect present, but it too appears to be an extremely consistent and powerful effect. The data compare favorably with those obtained through the use of simple stimuli (Sherif et al., 1958) and through the studies of contrast occurring as a result of the comparison of source-content (Hovland et al., 1957) variables in attitude formation.

The findings clearly indicate that the evaluation of the neutral concept is dependent upon the meaning abstracted in the conditioning of the critical concept. If degree of learning is used as a measure of strength of concept-attitude development (Rhine and Silun, 1958), it is apparent that the stronger of the two concepts (in this instance, the critical concept) is the anchor. The high correlation between the ratings of the critical concept and its corresponding neutral concept indicates that the greater changes occur at the extremes and the lesser changes (assimilation?) occur at the intermediate rating levels. Nevertheless, while the over-all contrast effects are strong, in no case do they equal in intensity the ratings of the critical concept-attitudes.

In the interpretation of the contrast effects a direct-opposite associative effect is highly favored in accounting for the evaluation of the neutral concepts subsequent to the training trials. This is consistent with Osgood's (1953, pp. 708–711) statement that opposites cannot be based upon semantic mediation processes, but are the result of highly overlearned verbal habits. Nevertheless, this does not preclude the possibility of a strong self-instruction habit, as a product of learning-how-to-learn, existing as a mediating process (see Campbell, Lewis, & Hunt, 1958, for example). Thus *S*s in a partially ambiguous situation behave as though a verbal rule such as that of "select-the-opposite" is particularly adaptive in an unstructured situation requiring evaluative responses. Once such a verbalization permits the structuring of a situation there follows the elicitation of the specific combination of opposites having the strongest habit strength within a hierarchy of possible responses. The particular hierarchy employed appears to be dependent upon a contextual definition based on the anchor point (Di Vesta & Bossart, 1958; Phillips, 1958).

There is one further implication of the present findings that concerns the design of investigations in this area. In the majority, if not all, studies on the conditioning of attitudes (Staats & Staats, 1957, 1958; Staats, Staats, & Biggs, 1958) and on the development of concept-attitudes (Rhine & Silun, 1958) the possibility of the contrast effect appears to have been neglected. Important in this respect was the fact that *S*s learned two concepts, concomitantly, one to positive evaluative words and the other to negative evaluative words, for example. In view of the present evidence, the results of said procedure would be to accentuate the stronger of the two evaluative poles but the rating of the weaker of the two concepts might well be the result of a direct (i.e., contrast effect) rather than mediated association. Under the circumstances it would appear that a study of two concepts that are learned concomitantly may be an inadequate procedure. Because of the possible interaction between the two, separate treatments in each group, rather than a covariance design, would be desirable.

On the basis of the Rhine and Silun (1958) study it would be expected that attitude development and attitude strength depend in large part on consistency of reinforcement (degree of learning). Provision in the present study was made for complete learning of the critical concepts by determining through pilot studies, the number of trials required for an abstraction to be made by *S*s. The analyses indicate that while differences in the mean number of errors between groups were to be found in the learning of the critical concepts and of the neutral concepts, the errors did not occur in a way that was related to the analyses performed. Thus, the differences in errors between those groups for which the ratings were directly compared (e.g., negative-evaluative vs. positive-evaluative) were in only one case (between the errors made in learning the neutral concepts learned respectively with the positive evaluative and negative evaluative critical concepts) significant. On the other hand, in every case the critical concepts were learned with fewer errors than the respective neutral concepts. Thus, some support is given to the hypothesis that essential contrast may be a superficial definition of the concept depending upon a direct, rather than mediated, association. Nevertheless, the phenomenon appears as a strong and potentially important influence in the learning of attitudes. The conditions of its demonstration in the present study calls attention to a heretofore neglected source of the occurrence of the "boomerang" effect in the development of attitudes.

SUMMARY

Each of four groups (N = 16 each) of *S*s learned one concept-attitude based on words strongly representing the evaluative or potency dimension of meaning and another concept-attitude based on words having neutral meanings on the respective dimension being manipulated. A control group ($N=10$) learned two concept-attitudes, both with neutral meanings.

Concept-attitudes are shown to be developed specific to the meaning dimension involved in the words used in conditioning. Furthermore, in support of the major hypothesis, concept-attitudes learned to neutral meanings, although conditioned to exactly the same words in all groups, were found to be rated by *S*s in direct opposition to the respective critical concept.

A mediation hypothesis was suggested as the best explanation of the changes in the critical concept. The contrast effect was explained in terms of the critical concept providing an anchor to which a direct-opposite response was applied for the rating of the less structured neutral concept. It was indicated that in previous studies of the verbal conditioning of attitudes some of the effect attributed to the mediation hypothesis may be little more than the formation of essential contrast. Note was also made that the contrast effect determined through contextual definition (while holding communicator effects constant) provides still another explanation of the so-called "boomerang" effect in the development of attitudes.

REFERENCES

CAMPBELL, D. T., LEWIS, N. A., & HUNT, W. A. Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *J. exp. Psychol.*, 1958, **55**, 220–228.

DI VESTA, F. J., & BOSSART, P. The effect of sets induced by labelling on the modification of attitudes. *J. Pers.*, 1958, **26**, 378–387.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

HARVEY, O. J., & CALDWELL, D. F. Assimilation and contrast phenomena in response to environmental variation. *J. Pers.*, 1959, **27**, 125–135.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PHILLIPS, L. W. Mediated verbal similarity as a determinant of the generalization of a conditioned GSR. *J. exp. Psychol.*, 1958, **55**, 56–62.

RHINE, R. J. A concept-formation approach to attitude acquisition. *Psychol. Rev.*, 1958, **65**, 362–370.

RHINE, R. J., & SILUN, B. A. Acquisition and change of a concept attitude as a function of consistency of reinforcement. *J. exp. Psychol.*, 1958, **55**, 524–529.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

STAATS, A. W., & STAATS, C. K. Attitudes established by classical conditioning. *J. abnorm. soc. Psychol.*, 1958, **57**, 37–40.

STAATS, A. W., STAATS, C. K., & BIGGS, D. A. Meaning of verbal stimuli changed by conditioning. *Amer. J. Psychol.*, 1958, **71**, 429–431.

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54**, 74–80.

(Received September 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 545–551

THE SPACING OF SEQUENTIALLY DEPENDENT TRIALS IN PROBABILITY LEARNING[1]

JEAN P. CHAPMAN[2]

Northwestern University

The present study is an investigation of some implications of the statistical model developed by Estes (1955) and Burke and Estes (1957) and a recent extension of this model by Witte (1959). More specifically it is a study of the effect of massed and distributed trials on response probabilities when the reinforcing events are sequentially dependent. Markov series of two-choice events are used in a Humphrey's-type verbal conditioning situation.

Recently there have been several studies using a sequentially dependent reinforcement schedule. Hake and Hyman (1953), investigating a different problem, studied both independent and dependent series of experimental events. Engler (1958) tested some predictions from the Bush-Mosteller model, using both independent and dependent series. She found that asymptotic marginal response probabilities approximately match marginal stimulus probabilities, but that asymptotic conditional response probabilities appear to exceed conditional stimulus probabilities except when the latter are extremely low.

A recent investigation by Witte (1959) used an extension of the Burke and Estes theory. To this theory Witte added a stimulus trace hypothesis to predict asymptotic behavior when trial spacing is distributed and the reinforcements are sequentially dependent. Witte predicted that conditional response probabilities should match the sequential dependencies between rewards only when trial spacing is massed. When the trial spacing is sufficiently distributed, these probabilities should approach the marginal probability for most marginal and conditional probability specifications. Witte tested this using rats in a T maze and found that the terminal conditional response probabilities for *S*s at each intertrial interval differed significantly in the direction predicted by the theory. However, the theory did not adequately describe the response probability at all points on the trial spacing continuum.

Witte presented a formula (his Equation 4) for the probability $p_{11}(n+1,t)$ of an A_1 response after trial n, given an E_1 occurrence on trial n and an intertrial interval of t. The solution of his Equation 4 is:

$$p_{11}(n+1,t) = \frac{1}{N_c + N_1(t)}\{N_c[(1-\theta)^n(F_c(1)-\bar{\pi}) + \theta(1-\bar{\pi})] + N_1(t)[(1-\theta\pi_{11})(1-\theta\bar{\pi})^{n-1}(F_1(1)-\pi_{11}) + \theta\pi_{11}(1-\pi_{11}) + \pi_{11}]\}, \quad [1]$$

with

$$Pr(E_1) = \bar{\pi} = \frac{\pi_{21}}{1-\pi_{11}+\pi_{21}} \quad [2]$$

---

[1] This paper is based on a PhD dissertation at Northwestern University. The author is indebted to John Cotton for invaluable assistance and advice in directing the dissertation, and to William K. Estes for his suggestions on the theoretical portion of this study.

[2] Now at University of Kentucky.

Thus the conditional response probabilities tend asymptotically to the corresponding conditional stimulus probabilities plus or minus "correction terms" of magnitude smaller than θ, and marginal response probabilities tend asymptotically to marginal stimulus probabilities plus or minus "correction terms" of magnitude smaller than θ.

Inspection of this formula reveals that it is a kind of weighted average of the Burke and Estes model for massed and distributed trials. Comparison of the Burke and Estes formulation with that of Witte reveals that the two models make a different prediction for massed trials except in the case of N_c/N_1 equal to zero. In order for Witte's model to be applicable to both massed and distributed trials this ratio must be greater than zero.

In the present investigation a comparison of these two models was made for data obtained with massed trials. A recent study by Anderson (1960), published since the completion of this work, tested the Burke and Estes (1957) model using data from a two-choice probabilistic task with sequentially dependent stimulus events. He found that the Burke and Estes model predicted terminal response levels quite well for the larger values of π_{11} but did poorly for the very low values.

Inspection of the model by Witte shows an important characteristic of this model. Each trial is affected only by what has occurred on the previous trial and by the length of the intertrial interval. This leads to the prediction that if, after a long period of massed trials spaced trials were given, the massed trials would have no effect on the spaced trials after the first spaced trial, and the response probabilities would shift immediately to the terminal values for the spaced trials. For longer spacing intervals the terminal response probabilities should be (for most marginal and conditional probability specifications) closer to the marginal probabilities plus or minus the correction terms.

As a test of this prediction *S*s were given a long series of massed trials with the reinforcements sequentially dependent. After the massed trials, each *S* was given a series of distributed trials with the same probability specifications as the massed trials.

METHOD

Apparatus.—A black board 39 in. × 31 in. separated *S*'s apparatus from the apparatus visible to *E*. On *S*'s side of the screen were two brown boxes 4 in. × 4 in. × 2 in. mounted 4 in. apart. On the vertical front face of each box was mounted a red light 1 in. in diameter and below it a toggle switch. From these boxes wires ran under the screen to the apparatus on *E*'s side of the screen. On this side of the screen were mounted a toggle switch which started a buzzer, two 15-w. bulbs each of which was lighted by the toggle switch on the appropriate box on *S*'s side of the screen, and two toggle switches each of which lighted the red light on the appropriate box on *S*'s side of the screen. The toggle switch on each of *S*'s boxes turned off the buzzer as well as turned on the appropriate light on *E*'s side of the screen. The two boxes in front of *S* and the corresponding lights were arbitrarily labeled E_1 and E_2. The *S* was not aware of this labeling. The E_1 box occurred on the right for half the *S*s and on the left for the other half.

Subjects.—The *S*s were 120 students from courses at the University of Chicago and the University of Kentucky. They were assigned to the 12 groups using a sequence from a random numbers table, with the restriction that 10 *S*s occur in each group. None of the *S*s had previously served in an experiment of this type.

Procedure.—Each *S* was given 500 trials. The first 480 trials for all *S*s were massed (one every 5 sec.) and were run without a break. There were four sets of marginal and conditional probability specifications used as shown in Table 1. Thirty *S*s were contained in each of the four groups corresponding to the four probability specifications. After the 480 massed trials, each of the four groups was divided into three subgroups. These three subgroups were given the 20 additional trials with intertrial intervals of 8 sec., 1 min., and 5 min., respectively. These intertrial intervals were from the presentation of the reinforcement until the sound of the buzzer starting a new trial, and were timed with a stop watch. The probability specification was the same for a given *S* throughout all 500 trials.

The *E* signaled the beginning of each trial by sounding a buzzer. The *S* then made his choice by pressing the toggle switch on the appropriate box. This stopped the buzzer and lighted the bulb which indicated his choice to *E*. Then *E* indicated the correct choice to *S* by pressing the toggle switch which lighted the light on the appropriate box in front of *S*.

The orders of presentation of the reinforcements were constructed to comply with four sets of marginal and conditional probability specifications. These probability specifications were chosen to enable comparison of the obtained results with those of Engler (1958). She had four groups that had the same conditional probability specifications as did the present study (i.e., $\pi_{11} = .70$ and .25) but had different marginal stimulus probabilities ($\bar{\pi} = .25$ and .50). A different sequence was made up for each *S* using tables of random numbers (Rand Corporation, 1955; Snedecor, 1946; Walker & Lev, 1953) each sequence being constructed according to the probability specifications for the group to which the *S* was assigned.

The theory specifies that these sequences be truly random. In order to comply with this specification the sequences were constructed in the following manner. The experimental event occurring on Trial 1 was determined by using a table of random numbers taking into account the value of π_1. The experimental event occurring on Trial 2 was then determined by using a table of random numbers taking into account the value of π_{i1}, $i = 1, 2$ where E_1 occurred on Trial 1, and so forth for all n trials. The resulting sequences do not agree exactly with the specified probabilities, but the group of sequences represents a random sample from the population of such sequences.

RESULTS

The learning parameter θ was estimated from the marginal data using the following equation:

$$\theta = \frac{\pi_1 - p_1(1)}{v(\pi_1) - \sum_{n=1}^{v} p_1(n+1)} \qquad [3]$$

where v is the number of trials used to estimate θ. Theta was estimated for each group, and the obtained values were .037, .062, .028, and .024 for the four groups, respectively.

TABLE 1

MARGINAL AND CONDITIONAL PROBABILITY SPECIFICATIONS FOR REINFORCEMENT SEQUENCES AND OBSERVED ASYMPTOTIC MARGINAL RESPONSE PROBABILITIES

Group	$Pr(E_1)$	$Pr(E_1 \mid E_1)$	$Pr(E_2)$	$Pr(E_2 \mid E_2)$	Observed Asymptote
I	.20	.70	.80	.9250	.18
II	.20	.25	.80	.8125	.13
III	.35	.70	.65	.8384	.33
IV	.35	.25	.65	.5962	.32

Massed data.—The obtained response probability curves are shown in Fig. 1. The proportions of A_1 responses following an E_1 occurrence are plotted for blocks of 20 trials. These curves are negatively accelerated and nearly asymptotic in the later trials, and they are quite variable, particularly in the later stages of practice.

The sections of Fig. 1 also contain the predicted response curves from the two theoretical models, that of Burke and Estes (1957) and that of Witte (1959). Comparison of the obtained curves with those predicted by Burke and Estes shows that the performance for Groups I and III is almost always well above the predicted curve. The terminal response values for these two groups also are above the predicted asymptotes. These groups both had reinforcement sequences constructed with a conditional probability (π_{11}) of .70.

Groups II and IV had reinforcement sequences with the conditional probability of an E_1 of .25, given an E_1 on the previous trial. The Burke and Estes (1957) predicted curve for Group II fits the data fairly well. This is true particularly in the early stages of practice, but there is some indication that the response proba-

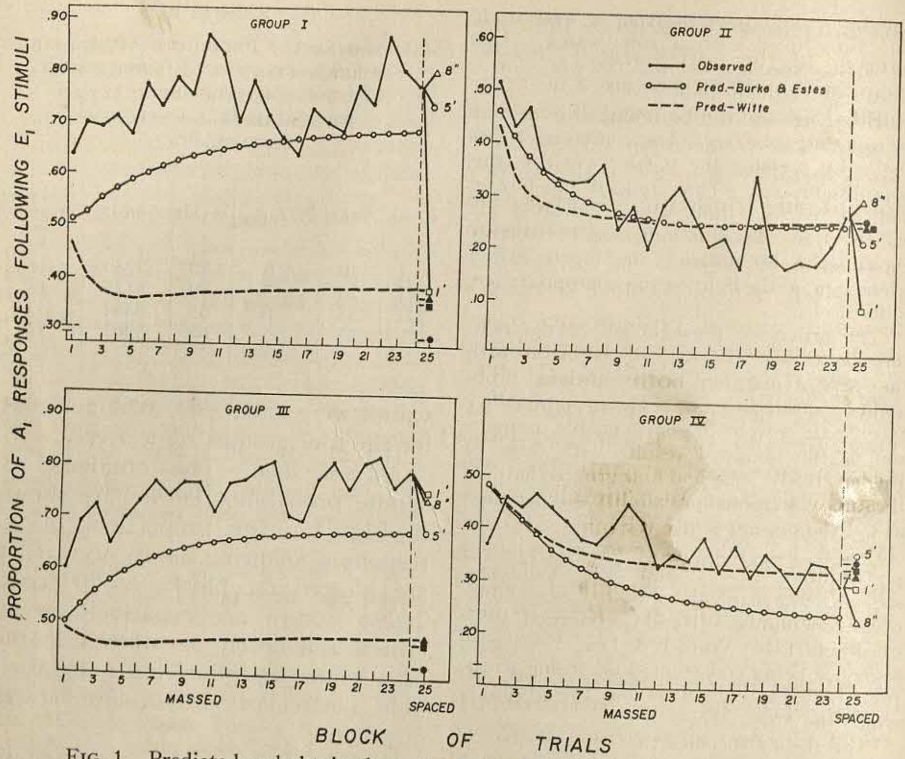


FIG. 1. Predicted and obtained conditional response probabilities for the four groups on massed and distributed trials.

bilities are becoming asymptotic beyond the predicted value. On the other hand the response curve for Group IV appears to be stabilizing at a value less than the predicted value.

A comparison of the Witte (1959) predictions with the obtained curves shows that Group IV produced the data that best matched the predicted curve. This group was the one used to estimate the necessary constants, N_c/N_1 and j_1.[3] In fact it was the only group which made possible such estimation. (As described earlier this model is a kind of weighted average of the Burke and Estes [1957] formulation for massed and completely distributed trials. In order for the Witte model to be used, it was necessary for the obtained terminal response data to be either equal to the predicted asymptote for the Burke and Estes massed trial model, or between the asymptotes for the massed and distributed models. This condition was met only in Group IV.) The Witte prediction seems to do about as well as the Burke and Estes prediction for Group II, particularly in the later stages of practice.

Figure 1 shows the N_c/N_1 ratio estimated from Group IV to be useless in predicting the performance of Groups I and III. Making a new estimate of N_c/N_1 to make this model fit the obtained data from these groups results in a negative ratio. Theoretically this would say that there are more elements (N_1) available after an E_1 occurrence than are present in the entire set of N_1+N_c elements, or that there are a negative number of elements available in the S_c set. These outcomes do not have much intuitive appeal as the theory now stands, and the incorporation of them would mean a major revision of the theory.

---

[3] The method for estimating the constants may be found in the author's dissertation which has the same title as the present paper. The values were estimated using the asymptotic response probabilities. The obtained values were $N_c/N_1 = 1.91$ and $j_1 = .0034$.

TABLE 2

OBSERVED AND PREDICTED ASYMPTOTIC PROBABILITIES OF A_1 RESPONSES FOLLOWING E_1 STIMULI

Group	Massed Trials			Distributed Trials		
	Obs.	B&E Pred.	Witte Pred.	Interval	Obs.	Pred.
I	.79	.69	.39	Massed		.69
				8 sec.	.80*	.38
				1 min.	.38	.37
				5 min.	.74*	.30
				Completely spaced		.23
II	.20	.24	.25	Massed		.24
				8 sec.	.30	.25
				1 min.	.10*	.25
				5 min.	.22	.25
				Completely spaced		.25
III	.79	.68*	.48	Massed		.68
				8 sec.	.75	.47
				1 min.	.76*	.46
				5 min.	.68*	.42
				Completely spaced		.37
IV	.32	.25*	.33	Massed		.25
				8 sec.	.24	.33
				1 min.	.26	.33
				5 min.	.35	.35
				Completely spaced		.37

Note.—Massed data were tested against the Burke and Estes prediction.

* These observed values differ from predicted with $P < .05$.

The left half of Table 2 shows *S*'s terminal response probabilities based on the last 100 trials and the predicted asymptotes from both models. The difference between the obtained asymptotes and the Burke and Estes (1957) predicted asymptotes was tested using a *t* test for all groups. The differences for Groups III and IV were significant ($P < .01$) but the differences for Groups I and II were not significant ($P = .10$ and $P = .60$).

Distributed trial data.—The obtained data for the 20 distributed trials for the four groups is shown at the far right of each section of Fig. 1. The Witte theory predicts that after a series of massed trials the proportion of A_1 responses following E_1 stimuli should immediately switch to a point closer to the marginal probability specification, plus or minus a correction term. The prediction is somewhat different for Group II where the marginal and conditional probability specifications are close together. The predicted massed asymptote for this group is .238 and the completely distributed asymptote is .249, with the conditional probability equal to .25 and the marginal probability equal to .20. This switch in the effect of spacing is due to the magnitude of the correction term for this group.

The Witte predictions for the three intertrial intervals for each group are shown at the far right of each section of Fig. 1. The groups with the greatest predicted switch are I and III, the groups with the greatest discrepancy between π_{11} and $\bar{\pi}$. These groups show some tendency to switch to the predicted values but not immediately as is required by the theory. The deviations of these groups from the predicted values were tested using the Wilcoxon (1949) signed ranks test. The results of this test are shown in Table 2. The 8-sec. and 5-min. values are both significantly different from the predicted values for Group I. For Group III both the 1-min. and 5-min. intervals produced significant differences. Groups II and IV are a little harder to interpret since the change predicted by the theory is not very great.

TABLE 3

COMPARISON OF OBTAINED MARGINAL RESPONSE PROBABILITIES AT 240 TRIALS OF ENGLER'S DATA AND PRESENT DATA

Conditional Probability	Marginal Probability			
	.20 (Present)	.25 (Engler)	.35 (Present)	.50 (Engler)
.25	.179	.33	.324	.493
.70	.157	.22	.322	.498

In fact the predicted distributed trial values are close to those predicted for the massed trials. Inspection of the results for these two groups reveals that the obtained values are scattered about the predicted values. Only one of the six differences tested for these two groups is significant (Group II: 1 min.).

For each of the four groups the Kruskal-Wallis H test was used to test whether the three subgroups having different intertrial intervals differed from each other in response probability. The differences were not significant for any of the four Groups. Group I showed a trend toward significance ($H = 5.59$, $P \cong .07$) with the 1-min. group being the deviant one.

Each of the spaced groups was compared with its respective massed asymptote to test whether there was a significant change in conditional response probability when trial spacing was introduced. These differences were tested using a sign test. Only one of the 12 subgroups showed such an effect. The 1-min. condition for Group I showed a significant change from the massed to the spaced condition.

Marginal data.—The theory of Burke and Estes (1957) predicts the marginal response probability equal to $Pr(E_1)$. Table 1 shows the obtained and predicted asymptotic marginal response probabilities. The obtained asymptotes are based on the last 100 trials. This table shows the obtained marginal responses to be consistently below those predicted by the theory. In order to compare the present data more directly with these of Engler (1958) the response probability values at 240 trials were computed from the present data. These values at 240 trials are compared with Engler's findings in Table 3. At this stage of practice the present data match the marginal probability specifications about as well as does the Engler data.

As an indication of the effect of conditional probability specification of the series on marginal response behavior, a 2 × 2 analysis of variance of marginal response probabilities was computed. No difference was found in marginal responses due to conditional probability specification ($F = .205$, $df = 1/116$). The interaction of marginal and conditional probability specifications also was not significant ($F = 1.94$).

DISCUSSION

The results of this study show that with sequentially dependent massed trials, the asymptotic conditional response probability does not uniformly match the conditional stimulus probability. Consistent with the results of Engler (1958), in some series when the conditional probability was high ($\pi_{11} = .70$) and the trials massed, the conditional response probability of an A_1 response following an E_1 stimulus exceeded the series specification. In the groups where the series specification was low ($\pi_{11} = .25$) the groups approximately matched the predicted asymptotes.

This study was designed also to test some inferences from two models, one of Burke and Estes (1957) and a modification of this model by Witte (1959). The results confirmed the finding by Anderson (1960) that the mathematical predictions showed some inadequacies in predicting the performance of human

*S*s in a verbal conditioning situation with sequentially dependent reinforcements.

The mathematical predictions about the distributed trials were not upheld in the two groups for which a large shift was predicted. The Witte model predicted that if, after a series of massed trials, spaced trials were given, the response probabilities would immediately shift to the terminal values for the spaced trials, the amount of shift determined in part by the size of the difference between the marginal and conditional stimulus probabilities. This immediate shift was not found in the data, though there seemed to be some indication of a trend of *S*s' responses toward the predicted values.

There is a large amount of variability noted in the response curves, especially in the groups that had series with long sequences. This is the case particularly in later stages of practice when it is not expected. One could speculate that many *S*s "caught on" to the presence of long sequences early in learning and were becoming less motivated as the series progressed. Many *S*s reported that they were "getting pretty bored" towards the end of the series and weren't "paying much attention to the task." They reported trying various repetitive responses (e.g., pushing the switch 10 times on one side, then 10 on the other) to break up the monotony. The *E*, also, noted many *S*s dropping characteristic ways of responding and adopting different patterns of responding in the latter part of the series.

The analyses of the marginal data revealed that *S*s overshot the more frequent stimuli or undershot the less frequent stimuli in the later stages of practice. Comparison of the present data with Engler's (1958) data showed that as the number of trials increased, *S*s tended to go somewhat beyond the predicted number of marginal responses. It would seem that with sufficient trials and sequentially dependent reinforcements *S*s tend not to match the marginal stimulus probabilities. It was also found that there were no differences in marginal response behavior due to the conditional stimulus probability specification.

SUMMARY

This experiment on probability learning was designed to test several predictions from the Burke and Estes (1957) model and from an extension of this model by Witte (1959). A total of 120 *S*s were tested on a Humphrey's-type light guessing apparatus. The series of reinforcing events were sequentially dependent.

The results showed that neither model adequately predicted performance on massed trials when the conditional stimulus probability was high, although both models did somewhat better at predicting performance when the conditional probability was low.

A prediction from the Witte model on performance on a series of distributed trials after a long series of massed trials was not upheld. There is some evidence that the model predicts qualitative but not quantitative changes from distributed trials.

The marginal response behavior revealed that response probability was consistently below the stimulus marginal probability.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–146.

ENGLER, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

RAND CORPORATION. *A million random digits.* Glencoe, Ill.: Free Press, 1955.

SNEDECOR, G. W. *Statistical methods.* (4th ed.) Ames: Iowa State Coll. Press, 1946.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid, 1949.

WITTE, R. S. A stimulus-trace hypothesis for statistical learning theory. *J. exp. Psychol.*, 1959, **57**, 273–283.

(Received August 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 552–559

YOKED COMPARISONS OF INSTRUMENTAL AND CLASSICAL EYELID CONDITIONING[1]

JOHN W. MOORE AND I. GORMEZANO

Indiana University

The purpose of the present investigation was to compare the acquisition and extinction of CRs in classical and instrumental avoidance eyelid conditioning when the comparison was independent of the number and pattern of UCS occurrences. In addition, the study was concerned with assessing the effects of omitting versus delaying onset of the UCS on a fraction of trials, under the two conditioning procedures.

Eyelid conditioning studies by Logan (1951), Kimble, Mann, and Dufort (1955), and Hansche (1959), comparing classical and instrumental avoidance procedures have indicated a higher level of conditioning with the classical procedure. Logan (1959), and Kimble et al. (1955), have attributed the inferior level of conditioning of the instrumental procedure to its partial reinforcement schedule. Kimble et al., further suggested that the observed differences may reflect some nonassociative effects of the UCS (i.e., each UCS occurrence develops an increment of acquired drive) and not an associative or learning factor. Moreover, Spence (1958) assumes that drive level (*D*) is in part a function of the general level of emotionality resulting from previous administrations of a noxious UCS. Since in instrumental avoidance conditioning the CR prevents the occurrence of the UCS, reinforcement theory would seem to imply, therefore, that classical conditioning is superior because of the greater number of reinforcements and higher motivational level of *S*. Stimulus-response contiguity theory (Guthrie, 1952) would also predict superior conditioning under the classical procedure presumably, however, because it assures the occurrence of a vigorous response on every trial. On the other hand, statistical learning theory (Estes, 1959)[2] would attribute inferior conditioning of the instrumental procedure to the greater variability of stimulus conditions. These interpretations, based on previously reported studies, have attributed differences in conditioning level to differences in the number and pattern of UCS occurrences. Presumably then, if an experimental comparison were made in which the pattern and number of UCS occurrences (and their possible nonassociative effects) were the same, these interpretations would lead one to expect no differences in performance under the two conditioning procedures. The present investigation attempted to make this comparison employing two reinforcement techniques: Depending on response contingencies, the UCS was omitted on a fraction of the acquisition trials or merely delayed by an interval longer than that shown to produce little or no conditioning (McAllister, 1953).

METHOD

Apparatus.—The *S*'s room was sound treated and contained two identical well

---

[1] This research was supported in part by a grant to the second author from the National Institute of Mental Health, Grant M-4034(A).

[2] W. K. Estes, personal communication, 1959.

illuminated enclosures permitting 2 *Ss* to be run simultaneously. The stimulating and recording components for the enclosures were identical. The enclosures were 3 ft. wide, 4 ft. deep, and 5 ft. high and their inside surfaces were painted a flat white. The *S* sat in the booth in a straight-backed chair and about 3 ft. in front of his eyes and slightly above his line of vision was a white rectangular box with a 10-cm. circular milk glass window. The CS consisted of an abrupt change in the illumination and hue (to reddish-orange) of this window upon discharge of a 2-w. General Electric NE 40 neon lamp. Also present in the enclosure was an intercommunication set and two 15-w. incandescent lamps which were kept lighted at all times to minimize beta responses (Grant, 1943; Grant & Norris, 1946). On top of the enclosures, and equidistant from their adjoining walls, rested an electric fan which provided a masking noise within the enclosures.

The *S* wore an elastic headband supporting a padded aluminum plate on which was mounted a rigid plastic stimulus air jet with a 2-mm. diameter orifice, and a microtorque potentiometer used to pick up *S*'s eyelid movement. Flexible plastic tubing from each of the air jets led to the two arms of a Y joint. The third arm of the Y joint was coupled to a rubber hose leading to *E*'s room and an ac-operated solenoid valve and tank of compressed nitrogen. The microtorque potentiometer was coupled to the right eyelid in a manner similar to that described by Grant, Schipper, and Ross (1952). A hypodermic tube was mechanically coupled to the shaft of the potentiometer and a length of piano wire was inserted in the tubing. To the other end of the piano wire was attached a small circular piece of thin aluminum. The aluminum tab was attached to *S*'s eyelid by adhesive tape, thus permitting movement of *S*'s eyelid to be transmitted to the shaft of the potentiometer. The frictional torque of the potentiometer was rated at about .004 in-oz. The signal from the potentiometer, generated by movement of the eyelid, was amplified by a Brush dc amplifier and recorded by means of an inkwriting Brush penmotor. Brush penmotors also recorded onset and termination of the CS and UCS.

Hunter-Brown and ATC electronic timers controlled durations of the CS, UCS, and CS-UCS interval. Activation of the timers and control of intertrial intervals was accomplished by a Western Union tape transmitter. The transmitter was pulsed by a cam-microswitch arrangement driven by a sychronous motor. The CS consisted of an abrupt change in illumination and hue of the milk glass window from .19 to .26 mL. which lasted for 600 or 3100 msec.[3] The UCS was a puff of compressed nitrogen of 100-msec. duration delivered to the right cornea and terminating with offset of the CS. The CS-UCS interval was either 500 or 3000 msec. The intensity of the air puff, measured at its point of delivery to the eye, was of sufficient intensity to support a 200-mm. column of mercury. The air puff was applied at a distance of $\frac{1}{2}$ in. from the right eye of *S*, and eyelid movements were recorded from this eye.

A thyratron circuit, when switched on, operated a relay whenever an eyeblink of a preset sufficient magnitude occurred. The system was designed so that eyeblinks occurring 200 to 500 msec. after CS onset, could prevent occurrence of the UCS or delay its onset for 3000 msec. A switch permitted *E* to determine whether CRs from one or the other of the two recording systems would activate the circuit.

Procedure.—All *Ss* received 70 acquisition and 20 extinction trials. The intertrial intervals were 15, 25, and 35 sec., in a random order with an average of 25 sec. Eyelid closures of at least 1-mm. deflection from the baseline in the latency range of 200 to 525 msec. during acquisition and 200 to 600 msec. during extinction were considered to be CRs.

Experimental design.—The *Ss* were randomly assigned to five groups of 24 *Ss* with sex differences and any variability associated with the two recording systems balanced among groups. Four of these groups constituted the cells of a 2 × 2 factorial design in which instrumental avoidance and classical reinforcement conditions were made orthogonal to two partial reinforcement procedures. For an avoidance conditioning group (A) and avoid-yoke group (AY) 2 *Ss*, 1 from each group, were run simultaneously in a yoked-chair procedure in which a sufficient eyelid response in the 200 to 500-msec. interval after CS onset, by the *S* in Group A, precluded the occurrence of the UCS for both *Ss* on that trial. If the *S* in Group A failed to make a CR, the UCS was presented to both *Ss* 500 msec. after CS onset. An avoid-delay group (AD) and avoid-delay-yoke group (ADY) received an analogous

---

[3] Calibrations were made with a Spectra Brightness Spot meter manufactured by the Photo Research Corporation, Burbank, California. Readings were taken with a filter corresponding to the "red eye" of the CIE luminosity curve.

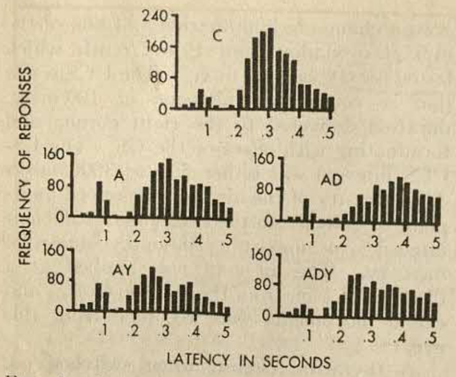


FIG. 1. Latency distributions of all eyelid responses in acquisition.

procedure in which a CR by the *S* in Group AD delayed the occurrence of the UCS for both *Ss* for 3000 msec. The fifth group, a conventional classical conditioning group (C), received the UCS on every trial at a CS-UCS interval of 500 msec. In extinction trials all groups received the CS alone for 600 msec.

Subjects.—The *Ss* were 120 men and women from introductory psychology courses at Indiana University. They were instructed to maintain a passive attitude toward the experiment and not to influence their natural reactions to the stimuli. Twenty-four *Ss* for each of the five conditions satisfactorily completed the experiment. In addition, 12 *Ss* were discarded from the experiment because of apparatus failure or experimental error.

RESULTS

The distribution of response latencies in acquisition for all five groups is shown in Fig. 1. The distributions are clearly bimodal with the first modes from the left showing the distribution of responses commonly taken as reflex responses to light (Grant, 1943). These responses are distributed about the 75-msec. mode with a latency range of approximately 40 to 120 msec. and occur with the highest frequency in Group A (avoid). The relatively low frequency of responses from approximately 100 to 200 msec., the beta range (Grant & Norris, 1946) suggests that precautions to maintain light adaptation were successful. Examination of the distribution of responses in the 200–525 msec. interval indicates some differences in the modal CRs among the five groups. The modal CRs for the two instrumental conditioning groups, A (avoid) and AD (avoid-delay), show slightly longer latencies than their yoked partial reinforcement groups, AY (avoid-yoke) and ADY (avoid-delay-yoke). Furthermore, Groups AD and ADY show a relatively greater frequency of CRs at the longer latency intervals. Inspection of records on delayed-UCS trials, however, indicated a relatively small number of responses in the 525–3000 msec. interval (not graphed). The mean number of eyeblinks in the 525–1025 msec. interval for Groups AD and ADY was 4.29 and 3.92, respectively, and in the 2500–3000 msec. interval, 8.04 and 5.25, respectively.

In Fig. 2 is shown the latency distribution of responses in extinction. The latencies of the modal CRs do not appear to have changed to any considerable extent from

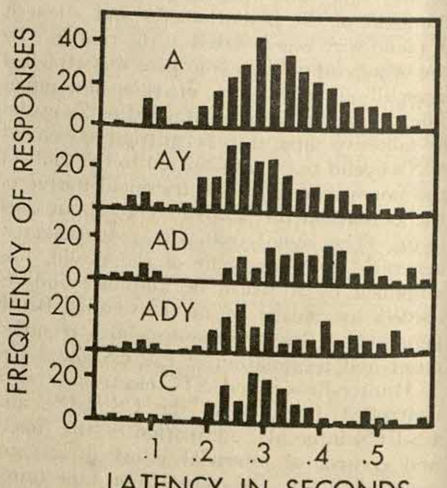


FIG. 2. Latency distributions of all eyelid responses in extinction.

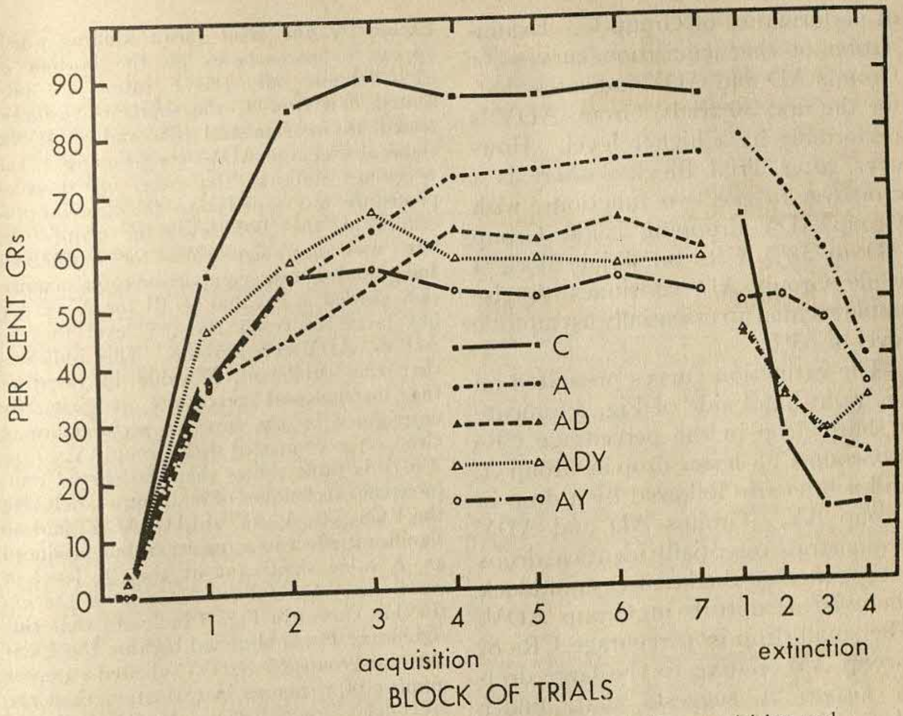


FIG. 3. The percentage CRs plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction.

acquisition to extinction. There are, however, relative to Groups C, AD, and ADY, a greater number of responses elicited in Groups A and AY.

Figure 3 presents the results of plotting percentage CRs in acquisition and extinction. The initial points for all acquisition curves are the percentage responses on the first trial. The remaining acquisition points are the percentage CRs for each 10-trial block and points on the extinction curves are plottings of 5-trial blocks. Examination of the figure indicates that all groups perform essentially alike on the first trial. The rate of acquisition of CRs by Group C then proceeds to surpass that of the other groups, reaching a maximum of 90% on Trial Block 3 and stabilizing at around 87% for the remaining blocks of trials. Groups A and AY pursue essentially identical courses of acqusition up to Trial Block 2, where performance is at about 55%. Group AY remains approximately at this level while Group A continues its rapid upward course to 72% CRs on Trial Block 3 and from there proceeds to show a very gradual rise to a level of about 76% on the last 10-trial block. Since the UCS fails to occur in Groups A and AY when CRs are made by *S*s in Group A, it is of interest to note that the stable performance of Group AY is being maintained under decreasing frequency of UCS occurrences. Also of interest, is the failure of Group A to reach asymptote. If further acquisition trials had been given it is conceivable that performance of this group might have reached the asymptotic level of performance of Group C. Examination of the acquisition curves for Groups AD and ADY indicates that, for the first 30 trials, Group ADY is performing at a higher level. However, after Trial Block 3 there is a crossover in the two functions, with Group ADY dropping below Group AD to 58% CRs on Trial Block 4 while Group AD continues to rise until reaching an essentially asymptotic level of 63%.

The extinction curves presented on the right-hand side of Fig. 3 indicate a sharp drop in the percentage CRs in Group C, a lesser drop in Group A, and a brief rise followed by a drop in Group AY. Groups AD and ADY demonstrate essentially identical drops except that on the last five-trial block there is an upturn in Group ADY. The small drop in percentage CRs of Group AY relative to the large drop in Group A suggests that, under identical pattern and number of UCS occurrences, classically conditioned CRs (AY) may be more resistant to extinction than instrumentally conditioned CRs (A). A comparison of the extinction trend of Group C with the four other groups, looked at in terms of frequency of reinforcements (i.e., presentation of the UCS 500 msec. after CS onset), reveals what would be expected from the previous literature on partial reinforcement.

The percentage CRs on Trials 41–70 of acquisition and 1–20 of extinction of the five groups presented in Fig. 3 were treated by the arc-sine transformation to bring about homogeneity of variance (Bartlett, 1947). The two instrumental conditioning groups (A and AD) and their yoked classical conditioning groups (AY and ADY) were then subjected to an analysis of variance in which response contingency (instrumental and classical) was orthogonal to partial reinforcement technique (omission and delay of UCS). Table 1 presents the analysis of variance for the acquisition and extinction data. Except for the error mean squares which appear in parentheses at the bottom of each column, only the *F* ratios are presented. In Row 1, the source of variation tested is instrumental (A and AD) vs. classical (AY and ADY) conditioning. The *F* values indicate that the instrumental procedure was superior to the classical procedure at the .005 level for acquisition, but was not significant for extinction. Individual *t* test comparisons for acquisition yielded a *P* value of .01 for A vs. AY but failed to reach the .05 level for the AD vs. ADY comparisons. This indicates that the significant *F* value in favor of the instrumental procedure is primarily determined by the superior performance of Group A as compared with Group AY. Row 2 of this table shows that the partial reinforcement techniques of omitting vs. delaying the UCS (i.e., A, AY vs. AD, ADY) had no significant effect in acquisition, but produced an *F* value significant at the .01 level in extinction. Careful examination of the extinction curves in Fig. 3 indicates that this significant *F* was obtained because the UCS-omission groups (A and AY) elicited a greater number of responses in extinction than the UCS-delay groups (AD and ADY). Row 3 presents the source of variation attributable to sex and Row 4 to differences in the two recording systems. The *F* values for these sources of variation failed to reach significance in acquisition or extinction. Row 5 presents the interaction effects of response contingency (instrumental and classical) and partial reinforcement technique (UCS-omission and UCS-delay). The *F* for acquisition was significant at the .05 level, but failed to reach significance in extinction. Examination of the acquisition curves in Fig. 3 shows that this significant interaction effect is attributable to the fact that under the instrumental procedure the UCS-omission group (A) was superior to the UCS-delay group (AD), but under the classical procedure the UCS-delay group (ADY) was superior to the UCS-omission group (AY). The remaining sources of variation failed to reach statistical significance in acquisition, but in extinction Rows 6 and 7 resulted in *F* values significant at the .05 level. The source of variation tested in Row 6 is the interaction effects of sex and response contingency. Examination of percentage CRs in extinction indicates that under the instrumental procedures men gave an average of 32.62% and women 53.09% CRs, whereas, under the classical procedures men gave 38.31% and women 34.84% CRs.

Row 7 presents the interaction effects of recording system and sex. The significant *F* is due to the fact that under one recording system men gave an average of 31.98% and women 51.01% CRs while under the other recording system men gave 38.95% and women 36.92% CRs. Interpretation of these significant *F* values (extinction for Rows 6 and 7) would appear to be trivial and probably represent Type I errors.

Statistical comparisons in acquisition were made of the conventional classical conditioning group (C) with Groups A, AY, AD, and ADY and indicated that Group C was significantly superior to all other groups (*t* tests, 46 *df*, $P < .01$ in all cases except Group C vs. Group A, where $P = .05$). Similar comparisons in extinction revealed that Group A made significantly more CRs during extinction than the four other groups ($P < .01$, except A versus AY where $P = .05$); all other possible comparisons failed to show significant differences.

DISCUSSION

The principal findings of this investigation were: (*a*) the instrumental avoidance conditioning group (A), when compared with a classical conditioning group equated for number and pattern of UCS occurrences, demonstrated a superior level of conditioning; (*b*) the partial reinforcement procedure of UCS-omission resulted in greater resistance to extinction than UCS-delay; (*c*) a significant interaction was obtained in acquisition between response contingency and partial reinforcement technique i.e., under the instrumental procedure the UCS-omission group (A) was superior to the UCS-delay group (AD), but under the classical procedure the UCS-delay group (ADY) was superior to the UCS-omission group (AY); (*d*) the conventional classical conditioning group (C) was superior to all groups in acquisition but demonstrated the least resistance to extinction, whereas, the avoidance group (A) showed the greatest resistance to extinction.

TABLE 1

ANALYSES OF VARIANCE OF PERCENTAGE CRs DURING ACQUISITION AND EXTINCTION

Source of Variation	*df*	*F* ratios	
		Acquisition	Extinction
Response contingency (RC): Instrumental (A and AD) vs. classical conditioning (AY and ADY)	1	9.05***	1.59
Partial reinforcement technique (PR): UCS-omission (A and AY) vs. UCS-delay (AD and ADY)	1	0.14	7.83**
Sex (S)	1	0.92	2.93
Recording system (RS)	1	3.70	0.51
RC × PR	1	5.09*	2.93
RC × S	1	1.15	5.81*
RS × S	1	3.77	4.49*
PR × S	1	1.06	1.87
PR × RS	1	0.46	0.06
RS × RC	1	0.00	0.01
RC × PR × S	1	1.83	1.71
RC × PR × RS	1	0.24	2.62
RC × S × RS	1	2.03	0.22
PR × S × RS	1	0.13	0.08
RC × PR × S × RS	1	0.45	1.61
Error	80	(383.61)	(592.18)

Note.—Number in parentheses are error mean squares. The analyses were performed on the arc-sine transforms of individual total percentage CRs for the last 30 acquisition trials and 20 extinction trials.

* $P = .05$.

** $P = .01$.

*** $P = .005$.

The first finding is contrary to the speculations of Logan (1951) and Kimble, Mann, and Dufort (1955), suggesting that the inferiority of instrumental to classical eyelid conditioning could be accounted for in terms of partial reinforcement. The acquisition data of the present investigation indicate, in fact, that instrumental eyelid conditioning when equated in terms of partial reinforcement pattern and number of UCS occurrences, is superior to classical conditioning. This difference can not be adequately interpreted by present versions of reinforcement theory, contiguity theory, or statistical learning theory which have limited their formulations to the role of the UCS as a determiner of performance, without taking into account the relationship of the CR and occurrence of the UCS under the two conditioning procedures. In the instrumental procedure the UCS occurs only in the absence of the CR and in the classical procedure it occurs independent of the CR. The superior performance of *S*s in the instrumental procedure appears to be due then, to their receiving the UCS when *needed*, i.e., when the CR fails to occur, whereas, no such consistent relationship exists for the classical partial reinforcement procedure. If one were to assume that the association between the CS and CR is formed in an all-or-none fashion, as has been suggested by Estes (1960), then presentation of the UCS would have a maximum effect on response probability when the CR fails to occur and no effect when the CR does occur. Since in the yoked classical procedure any number of trials may elapse in which neither the CR or UCS occur it would be expected that this relationship would lead to performance that is inferior to the instrumental procedure.

The second and third finding bears on the function of the delayed-UCS procedure in classical and instrumental eyelid conditioning. Reynolds (1958), Ross (1959), and Ross and Spence (1960), in investigations of classical partial reinforcement, have employed the delayed-UCS procedure to control presumed motivational factors on nonreinforced trials. These experiments did not compare delayed-UCS partial reinforcement with the usual partial reinforcement technique of omitting the UCS completely on a certain percentage of trials. In the present investigation there were no significant differences in acquisition between the UCS-delay and UCS-omission procedures but UCS-omission did lead to greater resistance to extinction. The third finding suggests, however, that the two reinforcement techniques might operate differently under classical and instrumental conditioning. Under the instrumental procedure the UCS-delay technique (relative to UCS-omission) appears to attenuate performance, whereas, under the classical procedure it enhances performance. The exact function of UCS-delay remains equivocal however, since the present comparisons of UCS-delay groups with UCS-omission groups are necessarily limited by the fact that their respective schedules of reinforcement were generated by the performance of the two instrumental groups (A and AD), and these schedules were not the same. (The percentage reinforcement of Groups A-AY and AD-ADY were 39.58 and 52.13, respectively.)

The fourth finding indicates that acquisition of CRs in classical eyelid conditioning on a 100% reinforcement schedule is superior to instrumental and classical partial reinforcement conditioning even when care is taken to control possible nonassociative effects of the UCS (i.e. presentation of the UCS at the delay interval of 3000 msec.). The greater resistance to extinction of the avoidance group (A), in which the UCS did not occur on every trial, can perhaps be best accounted for in terms of Sheffield's (1949) stimulus trace hypothesis of partial reinforcement. According to this formulation, the perseverative stimulus traces of nonreinforcement become part of the CS complex operating during the following trial. These perseverating cues are precisely those which occur in extinction. More indirect support for Sheffield's hypothesis is provided by the similarity in extinction of the 100% reinforcement group with the UCS-delay groups. Since the occurrence of the UCS on all conditioning trials for the UCS-delay groups made the stimulus after effects following reinforced and nonreinforced trials highly similar, the stimulus trace hypothesis would predict little or no differences in extinction. This finding in support of Sheffield's hypothesis has the logical restriction however, of becoming definitive only if the null hypothesis can be proved.

SUMMARY

The acquisition and extinction of CRs in classical and instrumental avoidance eyelid conditioning were compared when the comparison was independent of the number and pattern of UCS occurrences. In addition, the effects of omitting vs. delaying onset of the UCS on a fraction of trials were studied under the two conditioning procedures.

The results were as follows: (*a*) The instru-

mental avoidance conditioning group, when compared with a classical conditioning group equated for number and pattern of UCS occurrences, demonstrated a superior level of conditioning. (*b*) The partial reinforcement procedure of UCS-omission resulted in greater resistance to extinction than UCS-delay. (*c*) In acquisition the instrumental procedure of UCS-omission was superior to UCS-delay, but under the classical procedure UCS-delay was superior to UCS-omission. (*d*) A classical conditioning group on a 100% reinforcement schedule was superior to all groups in acquisition but demonstrated the least resistance to extinction, whereas, the instrumental avoidance group showed the greatest resistance to extinction.

These findings were discussed in terms of several contemporary hypotheses.

REFERENCES

BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

GRANT, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32**, 201–212.

GRANT, D. A., & NORRIS, E. B. Dark adaptation as a factor in the sensitization of the beta response of the eyelid to light. *J. exp. Psychol.*, 1946, **36**, 390–397.

GRANT, D. A., SCHIPPER, L. M., & ROSS, B. M. The effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

GUTHRIE, E. R. *The psychology of learning.* (Rev. ed.) New York: Harper, 1952.

HANSCHE, W. J. A comparison of operant, avoidance, and classical techniques in the conditioning of the eyelid response. Unpublished doctoral dissertation, University of Wisconsin, 1959.

KIMBLE, G. A., MANN, L. I., & DUFORT, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49**, 407–417.

LOGAN, F. A. A comparison of avoidance and nonavoidance eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 391–393.

MCALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–422.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

ROSS, L. E., & SPENCE, K. W. Eyelid conditioning performance under partial reinforcement as a function of UCS intensity. *J. exp. Psychol.*, 1960, **59**, 379–382.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

(Received August 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 560–564

THE INTERTRIAL INTERVAL IN CLASSICAL CONDITIONING[1]

WILLIAM F. PROKASY AND FRANCIS L. WHALEY

Pennsylvania State University

The purpose of the present paper is to examine more closely the adequacy of the construct of response-produced inhibition as an explanatory principle for the common observation that performance under distributed trial conditions is superior to that under massed trial conditions. The principle is most concisely stated as Hull's (1952) Postulate IX:

> Whenever a reaction (R) is evoked from an organism there is left an increment of primary negative drive (I_R) which inhibits to a degree according to its magnitude the reaction potential (${}_SE_R$) to that response. With the passage of time since its formation, I_R spontaneously dissipates approximately as a simple decay function of the time (t) elapsed . . . (p. 9).

One way in which this postulate can be tested is to compare the performance of independent groups of *S*s, each of which is conditioned with a different average intertrial interval. Three such studies (Baron, 1952; Prokasy, Grant, & Myers, 1958; Spence & Norris, 1950) have uniformly reported that response frequency in the classical eyeblink conditioning situation is positively related to the duration of intertrial interval.

An alternative test, employed in this study, can be effected by assessing response probability following individual intertrial intervals, with controls provided to balance out cumulative effects of the hypothesized inhibition. By training *S*s in a classical conditioning situation to frequency asymptote and employing extensive, balanced (but unpredictable) orders of intertrial intervals, it is possible to regroup the data of each *S* by separately evaluating response probability following each of the intervals employed. If Postulate IX is correct, response probability should increase as a function of the length of the specific interval separating trials.

METHOD

Apparatus.—The eyeblink conditioning apparatus employed was similar to that developed at the University of Iowa (Spence, 1953), as modified at the University of Wisconsin and described in Hansche and Grant (1960) and Prokasy, Grant, and Myers (1958). Eyelid movements were converted to changes in resistance by means of a piano wire attached to *S*'s right eyelid which in turn rotated the shaft of a microtorque potentiometer. Resistance changes from the potentiometer were electronically amplified and differentiated with both the response and its first time derivative recorded on a capillary pen ink-writing system with a paper drive operated at the rate of 150 mm/sec.

The *S*s were seated in a 3 × 5½ ft. room, facing a rectangular cubicle (34 × 34 × 28 in. deep) situated 31 in. above the floor. On the base of this cubicle facing *S* was a rectangular box, upon which were mounted two milk-white discs of 3 in. diameter. These were centered in the upper and lower halves of the CS box, with a signal light midway between them. Behind each milk-white disc was a 7-w. bulb, either of which could be employed as a CS. In this study only the upper light served in this capacity. The cubicle walls in *S*'s view were painted white, and were lighted by a 20-w. fluorescent lamp which was located out of *S*'s line of vision. This minimized the prospect of sensitized beta responses (Grant & Norris, 1947). The UCS, a puff of nitrogen, was delivered to the right cornea through a 5-mm. (inside diameter) glass jet.

The durations of CS and UCS signals were

---

[1] This research was supported by a National Science Foundation grant (G-7463) to the senior author.

controlled by Hunter-Brown interval timers. Trials were administered automatically by means of a taped program and a Western Union teletype transmitter. This permitted control of intertrial interval to within .25 sec.

Subjects.—The *S*s were 13 male students between the ages of 18 and 22 yr. They were volunteers from introductory psychology courses at Pennsylvania State University, and received course credit for their services. One of the 13 was terminated due to his failure to condition, apparently because of insensitivity to the puff stimulation.

Procedure.—Selected for investigation were the six intertrial intervals of 5, 10, 15, 20, 25, and 30 sec. These were chosen because prior data (Prokasy et al., 1958; Spence & Norris, 1950) indicated that performance differences were quite marked in the region of 15 to 30 sec.

Each *S* was given 97 reinforced trials in each of seven conditioning sessions. The first session was considered as a pretraining session, designed to bring *S* to asymptote. The remaining six sessions were test sessions, but data of only the final 78 trials in each of the test sessions were employed for analysis. By omitting the data of the first 19 test trials, the effects of any kind of recovery or regression were minimized. Thus, each *S* received a total of 679 trials, of which 468 were test trials employed for analysis.

In all sessions, including pretraining, all six intertrial intervals were administered in a predetermined, unpredictable order, with the restriction that there be a total of 16 trials following each interval length. (Note that this totals 96 intervals, which separate the 97 trials.) While the pretraining interval order was identical for all *S*s, a different order of intertrial intervals was administered for each of the six test sessions. Six different test-trial orders were employed, and were assigned to *S*s in a latin square design over the 6 test days. Thus, the overall design was a latin square with 2 *S*s in each of the six sequences of intertrial interval order. The only purpose in employing different sequences was the added assurance of being able to detect the existence, if any, of sequence-specific results.

The 78 test trials (i.e., those trials following the 78 test intervals) were ordered in such a way that each intertrial interval appeared 13 times, and that each of these intervals was preceded twice by all other intervals (including itself), during the session. This balancing was exercised in order to permit the assessment of possible residual effects from two trials preceding the test trial.[2] The net result was that, for each *S*, data pooled for the six test sessions provided a total of 78 trials immediately following each of the six intertrial intervals and 72 trials following each of the six intervals when they were two intervals prior to that trial.[3]

The CS was a brightness change of 3.8 to 190 mL. lasting 500 msec. The UCS, 50 msec. in duration, was directed at the right cornea from a distance of ½ to ¾ in., and flowed at the rate of 2 liters/min. (This rate was changed during the pretraining session for 1 *S* to a level of 2.75 liters/min. due to initial insensitivity.) The timed interstimulus interval was 470 msec., though the effective interstimulus interval was approximately 500 msec. This was determined by recording the transit time of the puff, approximately 30 msec. Because the six intertrial intervals were employed an equal number of times, the average intertrial interval was 17.5 sec.

Instructions were read to *S* at the beginning of the pretraining session. He was informed that he would experience two kinds of stimuli, a puff of air and a flash of light, and was instructed not to try to control his responses in any way. It was stressed that he should keep the CS panel in view at all times. Subsequent to the first day, instructions were not read, but *S*s were reminded, informally, of these instructions before each session. Due to difficulties in scheduling *S*s, it was not possible to run them on 7 consecutive days, though rarely was there a break between sessions of longer than 3 or 4 days.

Pen deflections of ½ mm. or greater were recorded as responses. While all responses

---

[2] The particular sequence is obtained easily by treating intertrial intervals as elements in a matrix. Form a symmetric 6 × 6 matrix with each row possessing all 6 elements. This will, when strung in a line, yield a system whereby each element is preceded by all other elements at least once. To insure that each element is preceded by itself requires the addition of, as a minimum, 3 elements prior to the appropriate rows, yielding a total of 39 elements. Perform the operation a second time and put the two strings together for a total of 78 elements.

[3] This differential comes about because certain intertrial intervals are employed in such a way that over the 6 test days there are six pairs of intertrial interval which constitute "extra" pairings. For purposes of analysis, the number of possible pairings had to be equal, hence, these "extras" were sampled out of the analysis.

TABLE 1

Frequency of Response for *S* DS as a Function of Test Day and Intertrial Interval

Test Day	Intertrial Interval					
	5	10	15	20	25	30
1	10	11	11	9	11	11
2	10	9	11	10	13	8
3	11	11	11	9	11	10
4	12	10	12	10	10	9
5	13	9	10	11	12	12
6	12	11	10	12	8	11

were recorded, only the first response in the latency interval of 200 to 500 msec. was counted as a CR. The time derivative of the response was used as a more precise index of response latency. To insure that apparatus cues were not available to *S*, a masking noise of 80 db., presented through earphones, was on continuously during each session.

Results

As a check for any unanticipated results inherent in test sequences over days, or for overall changes as a function of days, the total number of responses on each session was obtained for all *S*s and was employed as the dependent variable in an analysis of variance of the latin square design.

Based upon a residual variance of 484.11, the F ratios for day ($F = .1$; $df = 5/30$), sequence of intertrial interval orders ($F = 1.77$; $df = 5/30$), and intertrial interval order ($F = .07$; $df = 5/30$) were not significant. As anticipated, individual differences were significant ($F = 2.87$; $df = 6/30$). Because the overall 6 × 6 square was replicated, an assessment of square uniqueness was possible, and proved not to be significant ($F = 1.39$; $df = 20/30$). This result tends to confirm the assumption that those interactions, presumed to be negligible when latin square designs are employed, were negligible (Gourlay, 1955).

The basic analyses were performed on individual *S* data. In assessing the effects of intertrial interval on probability of response, a 6 × 6 matrix was formed. One dimension was sessions (confounded with specific intertrial interval order sequences) and the other dimension was intertrial interval. For purposes of illustration, a matrix of data for 1 *S* is provided in Table 1. An analysis of variance of response frequency as affected by sessions and intertrial interval was performed on such a matrix for each *S* separately. Because only a single value is present in each of the matrix cells, the error term employed was the interaction between session and intertrial interval. It was assumed, therefore, that this interaction effect was negligible, and to the extent that this assumption is not met the tests are conservative (Eisenhart, 1947).

The main effect of intertrial interval is significant for only 2 of the 12 *S*s. The data for all 12 are depicted in Fig. 1, those *S*s for whom the effect is significant being so indicated. Of immediate concern is the series of solid lines. Aside from two cases, the most striking feature is the zero slope of response frequency as a function of intertrial interval. This is a totally unanticipated result and it provides, insofar as negative results can be so interpreted, disconfirmation of the response produced inhibition hypothesis.

The results for the *S* (DW) whose data indicate a decreased response frequency as a function of intertrial interval are interpretable. This *S* had a tendency to doze momentarily between some trials during the last several minutes of the sessions. Subsequent examination of the records revealed that this was selective with the longer intervals. He was the only

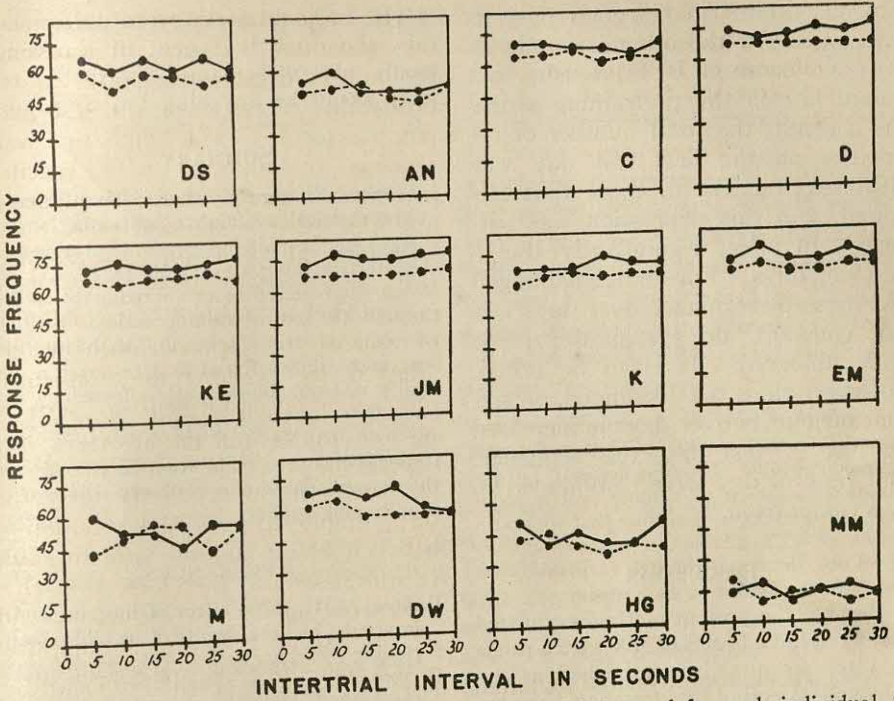


FIG. 1. Response frequency as a function of intertrial interval for each individual *S*. (Solid lines represent frequency following the immediate intertrial interval, dashed lines represent frequency on Trial *n* as a function of the interval preceding Trial *n*-1. Immediate intertrial interval effects are significant for *S*s M and DW.)

S in whom such behavior occurred. The second *S* (M) for whom the intertrial interval variable is significant was not consistent in his responding across this variable. We prefer to think that this is an instance of a Type I error.

The dashed line for each *S* in Fig. 1 depicts response frequency as a function of the intertrial interval *preceding* the interval immediately ahead of the test trial. Departure from slope zero would show residual aftereffects of the trial presented two trials before the test trial, and in no case was there significant departure.

An alternative index of cumulative effects was obtained by comparing response frequency following two successive 5-sec. intertrial intervals (5, 5) with response frequency following two successive 30-sec. intertrial intervals (30, 30). Of the 12 *S*s, 3 responded equally often following the 30, 30 and 5, 5 conditions, 1 responded more frequently following 30, 30, and 8 responded more frequently following 5, 5. Aside from ties, the likelihood by chance of 8 or more of 9 *S*s manifesting higher response frequencies to the same of two alternatives is .039 by a two tailed binomial test. This trend is in a direction *opposite* to that which would be predicted from the hypothesis of response-produced inhibition.

Analyses of response frequency as a function of days did not provide such unambiguous results. In 5 of the 12 *S*s, frequency varied significantly over days. In 2 cases, trends were manifest and explicable. In 1 of these, response frequency decreased systematically, accompanied by an obvious decline in responsiveness to the UCS. In the other case, the *S* was one for whom UCS intensity was raised late in the pretraining series. As a result, the total number of responses on the first test day was depressed relative to those that followed, and this depression was sufficient to affect significantly the *F* ratio of days. For the remaining 3 *S*s whose performance over days was not constant, no systematic trends were observed. It is to be noted, however, that the likelihood of significant effects over days is increased because *E*, test day, test sequence, and time of day were confounded in this comparison.

Discussion

Because the present method permitted a test of Hull's Postulate IX with a range of intervals already established as extensively affecting performance, the negative findings are incompatible with theoretical expectation. Insofar as the classical conditioning situation is concerned, the hypothesis that the massed-distributed effects observed in other studies are the result of response-produced inhibition is inadequate to account for present data.

At the empirical level, the data suggest that a kind of "range effect" may be operating. Prokasy and Chambliss (1960) compared fixed intertrial intervals with varied intertrial intervals and found that group performance did not vary as a function of the amount of variation in intertrial interval. Together with the present study, this indicates that performance varies as a function of the *average* intertrial interval, but that there are no immediate differential effects for particular intervals within the range employed. That there is evidence of cumulative effects in a direction opposite to that expected on the basis of Postulate IX suggests that there may be certain relatively local variations in response frequency which are dependent upon the spread of intertrial intervals employed. Should further research verify this, the causative agent of the commonly observed massed-distributed effects would be anything but apparent.

Summary

Twelve *S*s were given extensive training in the classical eyeblink conditioning situation, where the main independent variable was the intertrial intervals of 5, 10, 15, 20, 25, and 30 sec. interspersed in an unpredictable order through the entire training series. Pooling the data of trials following each interval separately, it was found that response probability was not differentially affected by the immediate intertrial interval. Such results are incompatible with the hypothesis that response-produced inhibition accounts for the massed-spaced effects observed in earlier conditioning studies.

References

Baron, M. R. The effect of long intertrial intervals on the limit of eyelid conditioning. *J. exp. Psychol.*, 1952, **44**, 438–441.

Eisenhart, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, **3**, 1–21.

Gourlay, N. *F*-test bias for experimental designs of the latin square type. *Psychometrika*, 1955, **20**, 273–287.

Grant, D. A., & Norris, E. B. Eyelid conditioning as influenced by the presence of sensitized beta responses. *J. exp. Psychol.*, 1947, **37**, 423–433.

Hansche, W. J., & Grant, D. A. Onset versus termination of a stimulus as the CS in eyelid conditioning. *J. exp. Psychol.*, 1960, **59**, 19–26.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Prokasy, W. F., & Chambliss, D. J. Temporal conditioning: Negative results. *Psychol. Rep.*, 1960, **7**, 539–542.

Prokasy, W. F., Grant, D. A., & Myers, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55**, 242–246.

Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.

(Received August 29, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 565–570

MEDIATION IN PAIRED-ASSOCIATE LEARNING

NAN E. McGEHEE AND RUDOLPH W. SCHULZ
Northwestern University *University of Iowa*

Russell and Storms (1955) have shown that language habits inferred from free-association norms are presumably capable of mediating the learning of verbal paired associates. Their *S*s learned to associate a word (B) with a nonsense syllable (A) in learning the first of two lists. The response unit B was the first word of an associative-chain (B-C-D) inferred from the norms. Following A-B learning, *S*s learned a test list consisting of A-D and A-X pairs. The A-D pairs were learned significantly faster than the A-X control pairs. However, as Russell and Storms (1955) hasten to point out, "The mere demonstration of mediational influences in learning . . . does not explain how the effect is achieved" (p. 292). One main purpose of the present experiments was to attempt to provide this explanation for the above situation.

As pointed out recently (Underwood & Schulz, 1960), it is analytically fruitful to conceive of verbal learning as a two-phase process. The first phase consists of a response-acquisition or response-recall phase where *S* is concerned with learning or recalling the response units per se. The second phase, the associative phase, involves learning to associate the response units with their appropriate stimulus units.

The usual interpretation of the Russell and Storms (1955) result has been the one that the associative phase of test-list acquisition was facilitated via the specific associative chains linking the A and D items of the respective pairs. An alternative interpretation, made apparent by the two-phase conception, is that facilitation occurred because the B items on the first list enhanced the availability of the D items during A-B learning by raising them in *S*'s response hierarchy. Moreover, this enhanced availability could facilitate test-list acquisition in the absence of any specific associative link between A and D. That is, the increased ease of response recall which would accompany heightened availability should facilitate test-list acquisition. The latter expectation is consistent with the results of a recent study in which it was found that the acquisition of paired-adjective lists was facilitated when the availability of the response units was enhanced by deliberate pretraining (Underwood, Runquist, & Schulz, 1959). The availability hypothesis was tested in Exp. I.

A second purpose of Exp. I was to compare performance under Russell and Storms' (1955) mediated and nonmediated conditions with performance under a condition in which first and test lists were unrelated (i.e., the practice control of transfer experiments). This was done to determine, in somewhat more absolute terms, the amount of facilitation which results from mediation in this situation.

The third purpose of Exp. I was to extend the generality of the Russell and Storms' (1955) results by replicating them with a design consisting of independent random groups and homogeneous lists as well as including *S*s of both sexes.

The results of Exp. I failed to

support the availability hypothesis. However, for reasons too lengthy to detail here, we remained skeptical as to whether the failure of the availability hypothesis necessarily implied that mediation had taken place via specific associative chains. Therefore, Exp. II was undertaken to investigate the matter further. Namely, if there is specific linkage, then a test list in which the A and D items are inappropriately paired (i.e., the analogue of the S_1-R_1, S_1-R_r paradigm of conventional transfer terminology) should be more difficult to learn than the nonmediated test list. Put another way, there should be mediated interference. Indeed, Norcross and Spiker (1958) have demonstrated just such mediated interference for associative links acquired entirely in the laboratory. In Exp. II an attempt was made to demonstrate mediated interference in the present situation.

A second purpose of Exp. II was that of determining the "criticalness" of the free-association norms as predictors of the mediation effects obtained in Exp. I. That is, while the B and D items were selected so as to minimize the possibility of a *direct* free-associative link between them, the method of selection *did not* preclude the possibility of a meaningful relationship between these items (e.g., Thief-Take, Wish-Need, etc.). Similarly, it did not prevent the occurrence of interlist relationships between response units in terms of formal similarity (e.g., Smell-Stem, Memory-Matter, etc.). Hence, if the interlist relationships between response units, along dimensions *not* predictable from free-association norms, were stronger in the mediated condition than in the nonmediated condition, then the interpretation of the superior performance in the mediated condition would require revision. Therefore, Exp. II included a condition in which the first and test lists were learned in reverse order (A-D, A-B). The logic of this arrangement is based on the assumption that the associative linkage defined by the free-association norms is unidirectional (i.e., B-C-D, but not D-C-B). Thus, making A-D the first list and A-B the test list, should reduce substantially, or even eliminate, the facilitation presumed to be mediated by habits inferred from free-association norms. Contrariwise, a relationship between B and D based on interlist response similarity should be unaffected by the reversal of the lists because it would be expected to be a bidirectional relationship. Storms (1958) has proposed and tested a similar hypothesis for a situation involving a single mediating term. He found clear evidence for bidirectionality with normatively unidirectional materials.

METHOD

Lists: Exp. I.—The relationships between first and test lists define the conditions of Exp. I, and are summarized in Table 1. The first-list designations shown in Table 1 will be used as abbreviations in subsequent references to the various conditions (e.g., the mediated condition will be called Cond. A2-B, the nonmediated condition will be Cond. A2-X, etc.).

The stimulus designations A1 and A2 are used to distinguish between the two sets of 10 nonsense syllables required by the design of Exp. I. The syllables were selected so that inter- and intralist similarity would be at a minimum. Their association values ranged from 0% to 27% according to Glaze (Underwood & Schulz, 1960, Appendix A). In minimizing similarity it was impossible, with one exception, to retain the syllables used by Russell and Storms.

The B, D, and X response units were the same words as those used by Russell and Storms (1955) and are shown in Table 3 of their article (p. 290). The details regarding the selection of these words may also be found there. In brief, C was the most frequent free-association response to B, and D the most

TABLE 1

RELATIONSHIPS BETWEEN FIRST AND TEST LISTS DEFINING CONDITIONS OF EXP. I

Condition	First List	Mediation Chain	Test List	Inferred Action
Mediated	A2-B	B-C-D	A2-D	$\begin{tikzcd} & B \arrow[r] & C \arrow[dr] & \\ A2 \arrow[ur] \arrow[rrr, dashed] & & & D \end{tikzcd}$
Nonmediated	A2-X	X-Y-Z	A2-D	$\begin{tikzcd} & X \arrow[r] & Y \arrow[r] & Z \\ A2 \arrow[ur] \arrow[rrr, dashed] & & & D \end{tikzcd}$
Availability	A1-B	B-C-D	A2-D	$\begin{tikzcd} & ? & \\ A2 \arrow[ur] \arrow[rr, dashed] & & D \end{tikzcd}$
Practice control	A1-X	X-Y-Z	A2-D	$\begin{tikzcd} & ? & \\ A2 \arrow[ur] \arrow[rr, dashed] & & D \end{tikzcd}$

frequent free-associate of C, but D was not among the 10 most frequent responses to B, e.g., B (Trouble), C (Bad), D (Good). The members of the X-Y-Z chain are related to one another in the same manner as the members of the B-C-D chain; however, the members of the two respective chains were not related, and must not be related, to each other. Each list consisted of 10 paired associates.

Conditions A2-B and A2-X correspond to Russell and Storms' (1955) chained (A-D) and unchained (A-X) conditions, respectively. The availability hypothesis would be supported if test-list performance in Cond. A1-B is superior to that in Cond. A1-X. Similarly, comparison of A2-B with A1-X will determine if absolute positive transfer resulted from mediation.

Lists: Exp. II.—Conditions A2-B and A2-X of Exp. 1 were replicated. The lists of Cond. A2-D were the same as those of Cond. A2-B in Table 1 except that A2-D was the first list and A2-B the test list. Condition A2-Br is most easily described as follows. Let A_1-B_1, A_2-B_2, etc. represent the syllable-word pairings of the first list. Similarly, B_1 is the beginning of the free-association chain B_1-C_1-D_1. The test list represents a random re-pairing of the respective A and D items: A_1-D_4, A_3-D_{10}, A_5-D_1, etc. This test list was List A2-D—the same one used in Exp. I and in Cond. A2-B and A2-X of this study.

Mediated interference will be demonstrated if performance in Cond. A2-Br is inferior to that in Cond. A2-X. If mediation in these materials is unidirectional, performance in Cond. A2-D and A2-X should not differ.

Procedure: Exp. I and II.—Each *S* was read standard instructions for paired-associate learning prior to learning the first list. This list was then learned to a criterion of three consecutive errorless recitations. Following the completion of first-list learning all *S*s rested for 4 min. After brief instructions to proceed as before, *S*s were presented the test list for 10 anticipation trials. For those *S*s not reaching a criterion of one errorless recitation during the first 10 trials, test-list acquisition was continued until this criterion was reached. The lists were presented on a memory drum at a 2:2-sec. rate with a 4-sec. intertrial interval. The 10 pairs in each list were presented in five random orders to prevent serial learning of the response units. The experimental session was limited to 50 min. for all *S*s.

Subjects.—The *S*s, Northwestern University undergraduates, were randomly assigned to conditions, with 30 and 24 *S*s per condition in Exp. I and II, respectively. The *S*s were naive with respect to the materials used, although most of them had served in other verbal learning experiments prior to their present service. When an *S* did not complete the experiment he was replaced by the next *S* appearing at the laboratory. There was no relationship between failure to complete the experiment and conditions.

RESULTS AND DISCUSSION

First-list acquisition: Exp. I and II.—Performance, in terms of mean number of trials to reach the criterion of three consecutive perfect recitations, did not differ significantly under the four respective conditions

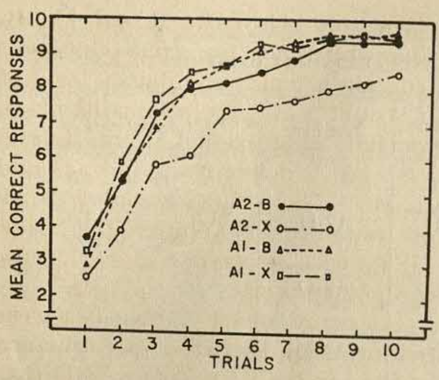


FIG. 1. Test-list performance in Exp. I as a function of various relationships between the stimulus and response units of the first and test lists. (See text for a complete description of these relationships.)

of either experiment (Exp. I: $F = 1.77$, $P > .10$; Exp. II: $F < 1$). In Exp. I the means were 18.33, 19.23, 18.90, and 15.43 for Cond. A2-B, A2-X, A1-B, and A1-X, respectively. The respective means for Cond. A2-B, A2-D, A2-X, and A2-Br of Exp. II were 13.62, 14.75, 14.71, and 15.87. The comparable difficulty of Lists A2-B and A2-D as first lists permits direct comparisons of test-list performance under Cond. A2-D with the other conditions of Exp. II, since the test list for this condition was List A2-B while List A2-D served as test list for the other conditions.

Test-list acquisition: *Exp. I.*—Test list performance is shown in Fig. 1. It is apparent from Fig. 1 that performance under Cond. A2-B, A1-B, and A1-X was essentially equivalent. Performance under Cond. A2-X was consistently inferior to performance under the other three conditions. The mean total number of correct responses during Trials 1–10 was 78.23, 65.43, 79.50, and 81.30 for Cond. A2-B, A2-X, A1-B, and A1-X, respectively. The standard error of these means ranged from 1.64 to 2.75. The only reliable ($P < .01$) differences among these means are those involving a comparison between Cond. A2-X and each of the other three conditions. The results for mean number of trials to reach the criterion of one perfect recitation were in complete agreement with those for total correct responses.

From these results it seems clear that Russell and Storms' (1955) findings were reliable and of some generality. The significantly ($t = 3.79$, $P < .01$) superior performance under mediated as contrasted with nonmediated conditions in the present study represents a reproduction of their results. The interlist relationship in these two conditions—S_1-R_1, S_1-R_2—is such that negative transfer would ordinarily be expected. Hence, the failure to find negative transfer under the mediated condition and the presence of a substantial amount of negative transfer in the nonmediated condition indicates that the inhibitory effects of the interlist relationship were somehow overcome in the mediated condition. However, comparison of performance under mediated and practice-control conditions makes it clear that the facilitation produced by mediation was not of sufficient magnitude to produce absolute positive transfer (see Fig. 1).

Finally, it is apparent that the avail-

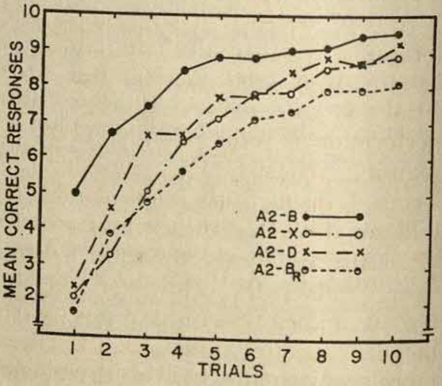


FIG. 2. Test-list performance in Exp. II as a function of various relationships between the stimulus and response units of the first and test lists. (See text for a complete description of these relationships.)

ability hypothesis was not supported by the results of the present study. That is, had the availability of the D units of the test list in Cond. A1-B been enhanced during first-list acquisition, then test-list performance under Cond. A1-B should have been superior to performance under Cond. A1-X. As can be seen from Fig. 1, it was not. Moreover, an analysis of overt errors failed to adduce any evidence indicating enhanced response availability in Cond. A1-B. Therefore, mediation effects, at least in the present situation, cannot be attributed to facilitation of the response-recall phase of test-list acquisition via enhanced response availability.

Test-list acquisition: *Exp. II.*—The performance on the test list for Trials 1–10 is shown in Fig. 2. The means for the total number of correct responses over the 10 trials were 83.04, 66.37, 71.62, and 61.54 for Cond. A2-B, A2-X, A2-D, and A2-Br, respectively. The standard error of these means ranged from 1.75 to 2.57. The overall differences among means are highly reliable ($F = 10.55$, $P < .01$).

Significant ($t = 5.12$, $P < .01$) superiority of performance in Cond. A2-B over Cond. A2-X again replicates Russell and Storms' (1955) results along with those of Exp. I.

As can be seen from Fig. 2, performance was initially comparable in Cond. A2-X and A2-Br, but after Trial 3 the curves diverge with performance under Cond. A2-X remaining consistently superior on Trials 4 through 10. The conditions did not differ significantly ($t = 1.45$, $P > .10$) in terms of total correct over Trials 1–10. However, the difference in mean performance is clearly in the expected direction. Furthermore, consistent with the divergence in the curves of Fig. 2 for these conditions, the means for the number of trials to reach the criterion of one perfect recitation were 8.92 for Cond. A2-X and 11.87 for Cond. A2-Br. This difference in trials to reach criterion is highly significant ($t=2.65$, $P < .01$). It appears justifiable to conclude that mediated interference has been demonstrated in Cond. A2-Br. This result agrees with the one obtained by Norcross and Spiker (1958) with associative chains learned in the laboratory.

Inspection of Fig. 2 reveals a trend toward slightly better performance, at least initially, in Cond. A2-D than in Cond. A2-X. However, when the means for total correct on the 10 test-list trials for Cond. A2-D and A2-X were compared they failed to differ significantly ($t=1.61$, $P>.10$). Indeed, in terms of mean trials to criterion Cond. A2-D was slightly inferior to Cond. A2-X. Furthermore, performance under Cond. A2-D was reliably poorer ($t=3.51$, $P<.01$) than under Cond. A2-B.

Thus interlist response similarity along dimensions not predictable from free-association norms does not appear to be a major contributor to the facilitation of test-list performance in the present situation. This result appears to conflict with the one described earlier, which Storms (1958) obtained. However, it should be recalled that his situation differed from the present one in that it involved only a single mediating term in contrast to the present two term mediating link. Perhaps, the potency of the "recency" effect to which he attributes his results varies inversely with the length of mediating chain. If this is the case, the present contradiction would be resolved easily.

In short, the results of Exp. I and II seem to provide fairly conclusive evidence that mediation in the Russell and Storms (1955) situation can be explained in terms of the facilitation of the associative phase of test-list acquisition via specific unidirectional associative chains linking the respective A and D items of the test-list pairs. Moreover, this interlist response relationship is associated with previously acquired language habits reflected by free-association norms.

SUMMARY

Two experiments were conducted to determine, among other things, how language habits inferred from free-association norms might mediate the learning of verbal paired associates in the Russell and Storms' (1955) situation. Their general procedure was replicated except that independent random groups and homogeneous lists as well as *S*s of both sexes were used. The interlist relationship between the stimulus syllables and response words of the first and test lists were appropriately varied to define the various conditions under which the 216 *S*s of the present experiments learned the 10-item lists on a memory drum at a 2:2-sec. rate.

From the results it was concluded that: (*a*) The response-recall phase of test-list acquisition is not facilitated in the mediated condition. (*b*) Mediated interference can be produced with the Russell and Storms' (1955) materials. (*c*) Under the mediated condition, the associative phase of test-list acquisition is facilitated via the specific associative chains linking the stimulus and response units of the respective pairs. (*d*) The free-association norms are "critical" in defining these associative chains. (*e*) Russell and Storms' (1955) results are reproducible and of considerable generality even though the facilitation produced by mediation does not result in absolute positive transfer.

REFERENCES

NORCROSS, K. J., & SPIKER, C. C. Effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, 55, 129–133.

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, 49, 287–293.

STORMS, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, 55, 390–395.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, 58, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received September 2, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 571–575

STUDIES OF DISTRIBUTED PRACTICE: XXI. EFFECT OF INTERFERENCE FROM LANGUAGE HABITS[1]

BENTON J. UNDERWOOD AND RUDOLPH W. SCHULZ
Northwestern University *State University of Iowa*

The last report in this series (Underwood & Schulz, 1961) indicated that when distributed practice facilitates acquisition of paired-associate lists it does so as a consequence of interference occurring during the response-integration phase of learning. The interference was produced by having *S*s learn lists containing elements which subsequently interfered with the response learning of the critical or test list. The present experiments continue the exploration of the interaction between intertrial interval and interference in verbal learning. However, the interference introduced into the lists results primarily from language habits brought to the laboratory by *S*s. Two types of experiments will be reported, distinguished by the locus of the interference. In the first type, interference is localized in the response terms and is presumed to hinder the acquisition of responses as such. In the second type, familiar words are used so that there should be no problem of response integration. Rather, the interference comes about as *S* attempts to attach particular responses to particular stimuli. If the previous findings are generalizable, facilitation by distributed practice may be expected with some level of interference as produced by the first type of experiment, but no facilitation should be expected in the second type of experiment.

METHOD

Three experiments will be reported. The procedures common to all experiments are as follows: (*a*) For distributed practice (DP) the intertrial interval was introduced after each trial and was always 30 sec. in length. Symbol cancellation was used to fill this interval. For massed practice (MP) the intertrial interval was 4 sec. (*b*) All lists were paired-associate lists. The rate of presentation was 2 sec. for the stimulus alone (anticipation learning), and 2 sec. for the stimulus and response together. Four different orders of presenting the lists on the memory drum were employed. (*c*) In a given experiment, no *S* served in more than one condition. Most *S*s, all undergraduate students at Northwestern University, were naive to formal verbal-learning experiments; if not, *S* was used only if the materials learned previously could not, in our judgment, interfere with the materials to be learned in the present experiments. (*d*) For each experiment a randomized order of conditions was made up and *S*s were assigned to this order as they appeared at the laboratory. (*e*) All lists were presented for a constant number of trials with *S*s instructed to give as many correct responses as possible on each trial.

Experiment I.—Three sets of eight trigrams were made up such that the average associative connection between letters may be spoken of as low, medium, and high. The associative connection was determined from the Underwood-Schulz (1960) tables. Within each set of eight trigrams formal intralist similarity was zero in that no letters were repeated. These trigrams were used as response terms with single-digit numbers as stimuli. Since each list was presented under MP and under DP, six groups of *S*s were required. Each of these groups consisted of 28 *S*s; all were given 20 anticipation trials.

Experiment II.—Two sets of eight bigrams each were constructed so that the average associative connection between the two letters for one list was low, for the other, high. The bigrams having low integration were: EY, CF, XV, DS, QW, JH, MK, and RZ. The bigrams having high integration were: MN, RA, CO, QU, XY, ET, JK, and

---

[1] This work was done under Contract Nonr-1228(15), Project 154-057, between Northwestern University and the Office of Naval Research. We are grateful to Geoffrey Keppel, who supervised the gathering of the data.

TABLE 1

PAIRED-ASSOCIATE LISTS USED IN EXP. III

List 1	List 2	List 3	List 4	List 5
CRUISER-HEAD	HEAD-MEASLES	BOB-HEAD	BOB-MEASLES	BOB-NITROGEN
EMERALD-WALL	BUS-MUMPS	BILL-BUS	BILL-MUMPS	BILL-SPARROW
BLUE-COTTON	KNIFE-POLIO	JOE-KNIFE	JOE-POLIO	JOE-GERMANY
THEFT-BEE	JOHN-CANCER	JOHN-CANCER	JOHN-CANCER	JOHN-CANCER
DOCTOR-BUS	WALL-RUSSIA	COW-WALL	COW-RUSSIA	COW-BLUEJAY
COPPER-TABLE	DOG-FRANCE	DOG-FRANCE	DOG-FRANCE	DOG-FRANCE
DOG-FRANCE	TROUT-ENGLAND	CAT-TROUT	CAT-ENGLAND	CAT-HYDROGEN
GASOLINE-MAPLE	DAISY-GERMANY	HORSE-DAISY	HORSE-GERMANY	HORSE-POLIO
WALTZ-OXYGEN	COTTON-BLUEJAY	RABBI-COTTON	RABBI-BLUEJAY	RABBI-MEASLES
WINE-GEOLOGY	MAPLE-CANARY	BISHOP-MAPLE	BISHOP-CANARY	BISHOP-RUSSIA
TRUMPET-KNIFE	GEOLOGY-SPARROW	MINISTER-GEOLOGY	MINISTER-SPARROW	MINISTER-SULPHUR
APPLE-TROUT	PRIEST-ROBIN	PRIEST-ROBIN	PRIEST-ROBIN	PRIEST-ROBIN
HAT-SECOND	BEE-NITROGEN	RHUMBA-BEE	RHUMBA-NITROGEN	RHUMBA-MUMPS
JOHN-CANCER	TABLE-HYDROGEN	FOXTROT-TABLE	FOXTROT-HYDROGEN	FOXTROT-ENGLAND
FOOTBALL-DAISY	WALTZ-OXYGEN	WALTZ-OXYGEN	WALTZ-OXYGEN	WALTZ-OXYGEN
PRIEST-ROBIN	SECOND-SULPHUR	TANGO-SECOND	TANGO-SULPHUR	TANGO-CANARY

DI. Single-digit numbers were again used as stimulus terms and acquisition was carried for 15 trials. Four groups of 15 *S*s were employed (MP and DP for each list). Subsequently, a partial replication was carried out by adding 15 *S*s to the two groups assigned the low-integrated list.

Experiment III.—The above two experiments were concerned with the effects of DP with varying degrees of response integration. In Exp. III the responses were common words. These words were taken from the norms given by Cohen, Bousfield, and Whitmarsh (1957). Their *S*s (a total of 400) were given category names under instructions to write down the first four instances of the category that occurred to them. The tables presented by the investigators are in terms of the frequency with which each response was given. For example, to the category An Article of Furniture, 392 of the 400 *S*s gave CHAIR, 346 TABLE, and so on. The nature of the lists constructed from such materials can best be understood by examining the lists. These are presented in Table 1.

First, attention is directed to List 2. In this list there are four sets of response words, of four each. The sets are distinguished in terms of category words. That is, MEASLES, MUMPS, POLIO, and CANCER represent responses given to the category stimulus "A Disease." The words RUSSIA, FRANCE, ENGLAND, and GERMANY are responses to the stimulus category A Country. The four words representing each category are words given with high frequency to the category stimulus. The assumption made was that because these items are all connected to a common category name (a common mediator), or because they have high associative connection among themselves as a consequence of cultural usage or as a consequence of being associated with a common mediator, differences in learning might be expected as a consequence of certain manipulations of these clusters. With these notions in mind, an examination of the structure of each list may be made.

List 1. There are no clusters of words from the same category.

List 2. Four clusters of four items each occur on the response side, with no clusters among the stimulus terms.

List 3. The reverse of List 2. That is, there are four clusters among the stimulus terms and none among the response terms.

List 4. Four clusters on both the stimulus and response side and the pairing of stimulus and response terms is such that all items of a given cluster on the stimulus side are paired with all items of a given cluster on the response side; e.g., animals are always paired with countries.

List 5. Four clusters on both the stimulus and response sides, but the pairing is such that the four members of a cluster on the stimulus side are paired with items from four different clusters on the response side.

In presenting the lists the order of the pairs was essentially random and four different orders were used. Each list was presented for 15 trials. A total of 150 *S*s was employed, divided into 10 groups of 15 each. Five groups were given the lists by DP and five by MP.

RESULTS

Experiment I.—The response measure was the mean total number of correct responses given over 20 trials. The results are shown in Fig. 1. The statistical analysis shows that only the *F* for response integration was significant; this effect is largely produced by differences between the low and medium list. The *F* for MP vs. DP was less than 1. Clearly, there is no evidence in these data that DP facilitates acquisition at any level of integration of these trigram responses.

Experiment II.—The response measure used was the mean total number of correct responses over 15 trials. For the bigram list with low integration the mean number of total correct responses given over 15 trials under MP was 47.4, and under DP, 62.5. For the list with high integration the means were 88.5 for MP, and 87.3 for DP. The *F* for interaction is 4.75, with 3.96 needed for the 5% significance level for 1 and 86 *df*. The *t* between MP and DP for the list with low integration is 2.55 when based on the variance of these two groups only, the σ_{diff} being 5.93. The acquisition curves over the 15 trials showed that DP was superior to MP throughout. Finally, it will be remembered that to the original 15 *S*s assigned each group, 15 were added to the two groups learning the list of low integration. For the original two groups assigned to this list the difference between MP and DP in total correct responses was 15.8; for the two added groups the difference was 14.4. We conclude that DP facilitates acquisition of bigrams of low integration and does not facilitate a list of well-integrated bigrams.

The learning of lists with response trigrams of low integration was not facilitated by a DP interval of 30 sec.; the learning of bigram lists of low integration was facilitated. As will be noted later, we are inclined to believe that there is a critical interaction between degree of response integration and length of intertrial interval. However, it is possible that in the case of bigrams the frequency of second letters is the critical factor. Both lists of bigrams had the same first letters, but to obtain low associative connection between the two letters the second letter must usually be of low frequency and to obtain high associative connection the second letter must usually be of high frequency. It is conceivable, therefore, that frequency of second letters is the critical difference between the two lists and not degree of integration. An indirect test was made of this notion by an experiment in which only single letters were used as response terms, one list having letters of low frequency (Q, Z, J, X, K, V, B, Y) and another having letters of high frequency (A, T, N, S, O, I, R, E). Again using single-digit numbers as stimuli, the lists were presented under MP and under DP. For both lists the means were almost identical for MP and DP. It may be concluded that it is not simply the frequency of second letters of the bigrams which are critical for the facilitation by DP found with the

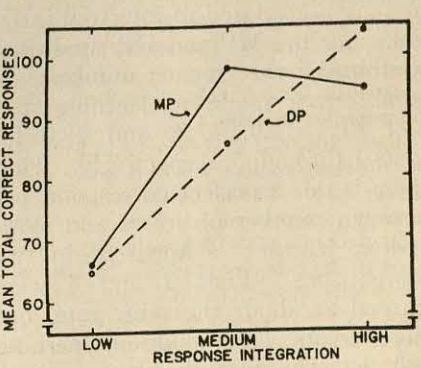


FIG. 1. Paired-associate learning as a function of degree of response integration of trigrams and massed vs. distributed practice.

TABLE 2

MEANS AND *SD*s OF TOTAL CORRECT RESPONSES IN LEARNING LISTS OF EXP. III AS A FUNCTION OF MASSED AND DISTRIBUTED PRACTICE

List	Massed Practice		Distributed Practice	
	Mean	*SD*	Mean	*SD*
1	197.40	22.70	197.33	24.18
2	189.80	29.83	175.27	32.35
3	176.13	35.74	187.60	21.85
4	157.33	35.47	143.40	36.24
5	159.20	24.27	147.73	28.96

list of low integration in Exp. II.

Experiment III.—The mean total numbers of correct responses given in 15 trials are shown in Table 2. An analysis of variance shows that variance attributable to list differences is highly significant statistically ($F = 13.20$, $df = 4/140$, $P < .01$). The F for MP-DP is 1.30. Not only does this fall far short of significance, but it may be noted that the direction of the difference in learning tends to to favor MP.

That differential interference is responsible for the differences in learning is suggested by the frequency of overt errors. These frequencies are, in general, inversely related to amount learned over the 15 trials and are not related to intertrial interval. For the MP and DP conditions combined, the mean numbers of total errors made in learning are: 8.8, 20.5, 14.9, 47.8, and 28.0 for Lists 1 through 5, respectively. That there is not a perfect correspondence between number of errors and slowness in learning is apparent in two comparisons. Lists 2 and 3 were learned at about the same rate, but more errors were made in learning List 2. The fact that more errors were made when the concepts were on the response side than when on the stimulus side corresponds to the results of earlier studies (Underwood, 1953) where clusters of highly similar adjectives were used. The second comparison is that between Lists 4 and 5. Again, the rate of learning was approximately the same for these two lists but the number of errors is much greater for List 4 than for List 5. All of these differences in error frequency maintain themselves even if adjustments are made for differences in opportunity to make errors because of differences in learning rate.

DISCUSSION

Considered in conjunction with previous experiments, the present three studies support two generalizations: (*a*) DP will facilitate acquisition only when interference occurs in integrating response terms, and (*b*) DP will not facilitate acquisition when interference, no matter how heavy, occurs in associating the stimulus and response terms of a paired-associate list. Obviously the first generalization must be limited in the sense that DP will not inevitably facilitate acquisition when interference occurs in acquiring response terms. In the present studies DP did not facilitate acquisition of trigrams at any level of integration; DP facilitated only the list of low-integrated bigrams. Clearly, a further narrowing of conditions under which DP will facilitate is required. Two statements will summarize our present position on this matter. First, some minimal amount of interference in response acquisition is necessary before DP will facilitate learning. Second, as amount of interference increases beyond this minimal amount, there will be a sharp interaction between length of the DP interval and the amount of interference. More specifically, it is believed that the greater the amount of interference the shorter must be the DP interval for facilitation to occur. The full elaboration of this position, along with some resolution of apparent contradictions to it now existing in the literature, has been made elsewhere (Underwood, 1961).

SUMMARY

Three studies evaluated the influence on learning of the interaction between intertrial interval and interference. The interference in Exp. I and II was centered in the response term of paired-associate lists. Distributed practice consisted of 30 sec. between each learning trial and massed practice 4 sec. between each trial. The results showed that with two-letter responses distributed practice facilitated learning when interference in learning the response was high, but not when it was low. Distributed practice did not facilitate the learning of lists in which the response terms consisted of three letters and this was true for all levels of interference.

Experiment III involved lists of 16 pairs of words in which interference was varied within lists by varying the stimulus-response relationships of clusters of words from the same concept categories. Although wide differences in learning occurred as a consequence of differential interference, no facilitation occurred by distributed practice.

The evidence indicates that distributed practice will facilitate acquisition only when interference obtains in response learning, and not when the interference occurs in stimulus-response learning.

REFERENCES

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. *Cultural norms of verbal items in 43 categories.* Technical Report No. 22, 1957, University of Connecticut, ONR Contract Nonr-631(00).

UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149.

UNDERWOOD, B. J. Ten years of massed practice on distributed practice. *Psychol. Rev.*, 1961, **68**, 229–247.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

UNDERWOOD, B. J., & SCHULZ, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. *J. exp. Psychol.*, 1961, **61**, 228–235.

(Received September 2, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 576–582

SEMANTIC SATIATION AMONG BILINGUALS[1]

LEON A. JAKOBOVITS AND WALLACE E. LAMBERT

McGill University

The phenomenon of bilingualism has been studied by linguists long before psychologists considered it of special interest. Weinreich (1953) in an exhaustive review of the literature on this topic reports several studies in which two forms of bilingualism have been identified. Following Osgood's (1952, 1953) development of a theory of meaning which uses a mediation principle as its main concept, Ervin and Osgood (1954) have formulated a psychological theory of bilingualism which incorporates the linguist's notion of two forms of bilingualism, accounting for the differences in types in terms of the manner in which bilinguals acquire and use their two languages. They define a "compound" bilingual as one having two functionally dependent linguistic systems, in contrast to the "coordinate" bilingual who makes use of two functionally independent language systems. The compounds are presumed to have a common mediation process for translated symbols in their two languages, while the coordinate bilinguals are believed to have relatively distinct mediation processes for translation—equivalent symbols in each of their languages. Lambert, Havelka, and Crosby (1958) have provided behavioral evidence in support of this theory.

Lambert and Jakobovits (1960) have recently established a convenient method for measuring a particular type of meaning change known to psychologists as "verbal satiation." This phenomenon can be defined as the decrease in the meaning of a linguistic symbol or sign as a result of its continued presentation. In their experiment, *S*s repeated a particular word continuously for 15 sec. and the amount of change in meaning was determined by comparing the intensity changes of the ratings of the words on various scales of the semantic differential (Osgood, Suci, & Tannenbaum, 1957). For example, the rating of the word FATHER on the 7-point "good-bad" scale would move, after repetition, from Position 1 on the scale ("extremely good") to Position 4 ("neutral or meaningless") when pre- and postrepetition ratings were compared. This phenomenon was viewed as a linguistic analogue of reactive inhibition (Hull, 1943). When a word is repeated the representational mediation process which constitutes the meaning of the word is continuously elicited. Theoretically, reactive inhibition (I_R) develops to the point where it suppresses or at least reduces the intensity of the representational process, resulting in the temporary loss or reduction of the meaning of the word. This reduction in meaning (verbal or semantic satiation) is indexed on the semantic differential as a decrease in the intensity with which the word is rated on the various scales. The size of the reduction is indicated by a "polarity-difference" score which is considered to be an index of "semantic satiability."

Semantic satiation becomes particularly relevant to the functional dependence or independence of the language systems of compound and

---

[1] This research was supported in part by the Canadian Defense Research Board Grant D77-9401-10.

coordinate bilinguals. When a compound bilingual continuously repeats a word, say HOUSE, the mediation process corresponding to that word should be suppressed, making for a loss of meaning of that word as well as the meaning of its translated equivalent, MAISON, since both words bear the same relation to the suppressed mediators. However, for the coordinate bilingual, translated "equivalents" function relatively independently so that the cross-linguistic satiation effect (hereafter called cross-satiation) should not be exhibited at all or, at least, should be exhibited to a lesser extent than in the contrasting compound case. Here, then, is a potential differentiator between the two types of bilinguals. The experiment which is described below has been specifically designed to test the validity of these theoretical speculations.

METHOD

Materials and Procedure

The method used here to measure semantic satiation was essentially the same as that described in a previous study (Lambert & Jakobovits, 1960). Briefly, this procedure consists of rating words on the semantic differential under two different conditions. In the first, *S*'s normal semantic profile for particular words is determined by having him rate them on a number of semantic (bipolar) scales. Then, in the second or treatment condition, *S* is required to repeat rapidly a word aloud for a period of 15 sec. just before he gives a rating for that word. By comparing the semantic ratings under the two conditions, the amount of decrease in the intensity of ratings attributable to repetition can be determined by means of the polarity-difference score.

The words (see Table 1) and scales were printed on 3 × 5 in. white index cards and placed in a Kardex folder in such a manner as to enable *E* to expose them in a predetermined randomized order. Five semantic scales were used in English for the English words and their French equivalents for the French words. These were: good-bad (bon-mauvais), pleasant-unpleasant (agréable-désagréable), large-small (grand-petit), strong-weak (fort-faible), active-passive (actif- passif). The *S* made his ratings by pointing with a stylus to one of the seven positions along each scale and his responses were recorded. The various treatment conditions which were used are described below and summarized in Table 1.

Normal semantic differential.—The experiment started with the determination of *S*'s semantic ratings under normal conditions for eight words each rated along five scales (making a total of 40 responses in random order). For example, *E* would expose the word FATHER which *S* pronounced aloud once; then *E* exposed a scale (e.g., good-bad) and *S* made his rating. This procedure was followed until all words were rated on all scales. These 40 ratings served as points of comparison for the various treatment conditions described below.

Satiation condition.—Four of the words rated above (two in English and two in French) were presented again (20 responses). The *E* first exposed a word which *S* repeated aloud for 15 sec. at a rate of about two repetitions per sec. Then *E* exposed a scale, a signal to *S* to stop repeating and make his response.

Cross-satiation condition.—Two other words among those rated under the initial normal condition (one in English and one in French) were presented again (making a total of 10 responses). The *E* first exposed a word such as KITCHEN and *S* repeated it aloud for 15 sec. Then *E* exposed a second word which was the translated equivalent word for the first, in this case CUISINE. The *S* pronounced this second word once, after which *E* immediately exposed a scale and *S* made his rating. The *S* had been previously instructed to "always make your response to the second word, not to the first." By comparing the ratings under the present condition with those made under the initial normal condition, the effect of the repetition of its translated equivalent upon the semantic ratings of a particular word (cross-satiation) was determined.

Different-word control condition.—Two other words among those rated under the initial normal condition were presented again. The *E* first exposed a word such as SKY which *S* repeated aloud for 15 sec. Then *E* exposed a second word in the *other* language (FLEUR) and *S* pronounced it once, after which *E* immediately exposed a scale and *S* made his response to the second word. This condition was therefore parallel to the cross-satiation condition except that the word repeated and the word rated did not bear the

TABLE 1

SCHEMATIC REPRESENTATION OF THE PROCEDURE

Condition	Description	Words Used	Examples[a]
Normal semantic differential	Ratings given under normal conditions.	FATHER, COP, VILLE, GARÇON, CUISINE, HOUSE, FLEUR, HATE	"FATHER": "HOUSE":
Satiation	15-sec. repetition of a word followed by rating of that word.	FATHER, COP, GARÇON, VILLE	(FATHER): "FATHER" (GARÇON): "GARÇON"
Cross-satiation	15-sec. repetition of a word followed by rating of its translated equivalent.	CUISINE, HOUSE	(KITCHEN): "CUISINE" (MAISON): "HOUSE"
Different-word control (diff. language)	15-sec. repetition of a word followed by rating of an unrelated word in the other language.	FLEUR, HATE	(SKY): "FLEUR" (BATEAU): "HATE"
Different-word control (same language)	15-sec. repetition of a word followed by rating of an unrelated word in the same language.	MONEY, VÉRITÉ	(SMOKE): "MONEY" (PAYSAN): "VÉRITÉ"

[a] In the examples, the words in parentheses are repeated for 15 sec. and the words in quotation marks are rated.

relationship of translated equivalents, but were semantically unrelated words. This condition was intended to serve as control for the cross-satiation condition and was introduced to determine the effect of the task of repetition per se upon the rating of another word.

It should be mentioned here that the compound group was recalled to the laboratory and tested a second time for reasons to be discussed below (see Discussion). During this second testing a different-word control condition was administered in which the word repeated (SMOKE) was in the *same* language as the word which was rated (MONEY).

The order of presentation of the various treatment conditions was systematically varied to control for and permit identification of possible order and fatigue effects.

Subjects

The *Ss* were a heterogeneous group of 62 English-French bilinguals (college students, graduate students, housewives) who were paid for their cooperation. An attempt was made to use "balanced" bilinguals by asking for volunteers who felt "equally proficient in both languages." However, to insure this requirement, Lambert's (1955) automaticity measure was administered to every *S*. This measure involves the use of eight finger keys, the stems of which are differently colored, an exposure apparatus which directs *S* to depress a particular key with directions appearing randomly in either language, and a chronoscope measuring latency of response. A significance test can be applied to test the hypothesis that balanced bilinguals will show a mean differential latency of response which is not significantly different from zero. Two bilinguals who were not balanced were excluded from analysis of results.

In order to determine compoundness or coordinateness among bilinguals, detailed information was obtained from each *S* concerning how, when, and where his languages were acquired. The variables which have been shown by Lambert et al. (1958) to be relevant were used as a basis for classification, i.e., whether *S* had learned the two languages in the same cultural setting or not, whether he has used both languages interchangeably inside and outside the home, etc. On the basis of this interview *Ss* were classified as either compound or coordinate, and their results analyzed separately. Separate groups of 30 compounds and 32 coordinates were thus obtained.

RESULTS

The initial polarity scores given in Table 2 were derived by totaling each *S*'s ratings along the 4-point intensity scale and determining a group mean. Thus 16.67 (cross-satiation) is a group mean calculated for 10 ratings per *S*. Since the initial polarity scores could range from 0 to 30, it will be noted that all of the actual scores fall near the median, ranging from 13.94 to 18.07. Consequently it is unlikely that these initial differences influenced the experimental treatment effects to be noted below.

Table 2 also presents the mean change scores which are the differences between initial and final polarity scores. For example the mean change for the compound group after the satiation treatment was −1.27 which indicates a change from an initial polarity score of 16.50 to a final polarity score of 15.23.

Tests of significance are given for departure from zero of each group's mean change, and for differences between mean changes for the compound and coordinate groups on each of the three conditions. (All t tests quoted in this experiment are two tailed.) It can be seen that the two groups differ significantly from each other under the cross-satiation condition but not under the other two conditions. This finding substantiates our prediction that compound and coordinate bilinguals will react differently to the cross-satiation treatment. We will now examine the results with respect to each group's behavior under the three different conditions.

Satiation.—The compound group has a negative mean which departs significantly from zero, whereas the coordinate group has a negative mean which fails to reach statistical significance. The result is ambiguous since there is no significant difference between the two groups.

Cross-satiation.—The significant negative mean for the compounds is as expected. The significant *positive* mean for the coordinates was not anticipated. According to the original hypothesis, it was predicted that coordinates will either satiate cross-linguistically to a lesser extent than compounds or will not be affected

TABLE 2

INITIAL SEMANTIC DIFFERENTIAL POLARITY AND MEAN CHANGE IN POLARITY OVER THE SUM OF TWO WORDS AND FIVE SCALES FOR THE COMPOUND AND COORDINATE GROUPS

Condition	Compounds $N = 30$					Coordinates $N = 32$					t[a]
	Initial Polarity		Mean Change			Initial Polarity		Mean Change			
	Mean	*SD*	Mean	*SD*	t[b]	Mean	*SD*	Mean	*SD*	t[b]	
Satiation[c]	16.50	3.90	−1.27	2.74	2.54**	15.03	4.96	−0.25	2.04	0.69	1.64
Cross-satiation	16.67	4.97	−3.27	3.23	5.45***	13.94	5.23	1.16	2.75	2.37**	5.74***
Different-word control	18.07	4.99	−1.30	3.42	2.06**	17.47	4.88	0.53	4.00	0.75	1.89

[a] t for average change in polarity between compounds and coordinates.

[b] t for departure from zero, two-tailed.

[c] Since there were two English and two French words under this condition, the mean of each pair was used to permit comparisons with the scores in the other conditions.

** $P < .05$.

*** $P < .01$.

by the cross-satiation treatment. It now appears that the coordinate group behaves in exactly the *opposite* manner from the compounds, exhibiting a "generation" of meaning instead of satiation.

Different-word control.—The coordinate group seems to exhibit no systematic effect or change under this control condition, whereas the compounds have a negative mean which departs significantly from zero. However, the difference between the two groups does not reach statistical significance.

A postexperimental interview revealed that none of the *S*s were aware of the purpose of the experiment. Most of them believed that *E* was investigating "the differences in my reactions in English and French." The *S*s who stated that their ratings changed (from normal to treatment conditions) thought this was due to the fact that "I forgot what I had done before," or that "I thought of something else the second time." Only 1 *S* stated that his ratings became less extreme under repetition conditions (including control) and attributed this effect to his "confusion." It appears, then, that the semantic satiation effect operates without explicit awareness on the part of *S* of the purposes of the experiment or of the fact that his ratings change in a systematic manner.

DISCUSSION

Although our main prediction has been borne out, namely that compounds cross-satiate, coordinates do not, it is apparent from the pattern of results in Table 2 that our original account of the effect of the three treatment conditions is incomplete and that our theoretical model requires extension. We shall formulate the following two hypotheses which suggest themselves from the present results, but which are also in agreement with some other findings to be mentioned below.

Hypothesis A.—Bilinguals as a group are less susceptible to the satiation effect than are monolinguals. In a pilot study with 27 bilinguals we have used essentially the same method and procedure as that described in the present investigation except that the automaticity measure was not administered and therefore no statements can be made as to their comparative proficiency in the two languages. The combined mean for the 27 bilinguals under the satiation condition was found to be 0.16 ($SD = 2.07$) which is not significantly different from zero ($t = 0.40$). In the present study, when we combine the scores of both compounds and coordinates under the satiation condition ($N = 62$) we obtain a negative mean of -0.76 ($SD = 2.46$) which is also not significantly different from zero ($t = 0.77$). These findings are in striking contrast with experiments on monolingual *S*s in which a significant satiation effect was repeatedly obtained (Lambert & Jakobovits, 1960).

Added support for this hypothesis comes from unpublished evidence of a slightly different kind. Carroll's (Carroll & Sapon, 1955, Test V) paired-associate test was administered to 18 *S*s used in another study (see Lambert & Jakobovits, 1960) and their scores on this test were correlated with their polarity-difference scores under the semantic satiation treatment. A product-moment correlation coefficient of -0.66 ($P < .01$) was obtained. This finding suggests that efficient verbal learning (of the paired-associate type) depends on an ability to resist the semantic satiation effect. Bilinguals, and especially balanced bilinguals, have demonstrated their skill in verbal learning and it is likely that they represent a sample of *S*s who are low on semantic satiability.

Hypothesis B.—An experimental situation in which the bilingual is required to switch from one language to the other is inefficient and inhibitory for the compound, facilitative for the coordinate.

This hypothesis seems to contradict the position taken by Ervin and Osgood

(1954) who state that "theoretically, decoding from a foreign language should be facilitative for the compound system, since different inputs are associated with the same representational reactions or meanings" (p. 143). It seems equally possible, however, that the facilitation provided by the similarities of the mediators is counteracted by the interference effects produced when the two languages are to be used separately, as in ordinary life situations. This is partly recognized by Ervin and Osgood (1954) in their statement that "interference is most likely to occur when the languages are closely related and the cultures and experiences associated with the languages are alike" (p. 141). It seems that in order to reach bilingual fluency, the compound must develop an inhibitory mechanism which would reduce the availability (and therefore the potential interference) of the second language while the first is being used. Then, if he is required to switch from one language to the other, he will behave less efficiently than when he is permitted to function with one or the other of his languages.

The cross-satiation and different-word control conditions used in the present experiment represent a language switching situation. As the compound bilingual repeats a word in one language (e.g., SKY), and is then presented with a word in the other language (FLEUR), the decoding of this second word will be inefficient, reflected as a systematic change toward less intense positions or toward the middle or meaningless point on the semantic scales. The coordinate bilingual, however, will behave more efficiently in the same situation since the two language systems are relatively independent and less subject to interlingual interference. It even appears that the switching process is facilitative, as indicated by the increased polarization of the ratings under language switching situations (i.e., different-word control and cross-satiation). In the case of the compounds, the cross-satiation treatment has a neutralizing effect on the meanings of translated equivalents, whereas for the coordinates the repetitive elicitation of mediators in one language appears to "liberate" the mediators of translated equivalents.

The following deductions can be made from the two hypotheses: (*a*) Since the cross-satiation treatment for the compounds involves both the inhibition of mediators as a result of their repeated elicitation and the inhibition of the decoding process of the second language, the negative mean under this condition should be greater than either the mean of the satiation condition or that of the different-word control in which only one effect is operating at a time. Correlated t tests between these three conditions revealed that the cross-satiation condition has a significantly greater negative mean than that of the satiation condition ($t = 2.25$, $P < .05$) as well as that of the different-word control ($t=3.28$, $P<.01$), thus supporting the first deduction. Furthermore, it appears that the satiation and different-word control treatments have equal effects since the two means are not significantly different ($t = 0.03$). (*b*) Since language switching is facilitative for the coordinate group, there should be a significant difference between both cross-satiation and different-word control conditions and the satiation treatment (which does not involve language switching). Correlated t tests show that while the difference between the cross-satiation and satiation treatments are significantly different in the predicted direction ($t=2.14$, $P<.05$), the difference between different-word control and satiation conditions does not reach an acceptable level of significance ($t = 1.18$, $P < .30$). The deduction is thus only partially confirmed. The cross-satiation and different-word control treatments seem to have equal effects ($t = 0.68$). (*c*) If a different-word control condition is administered to the compounds in which the word repeated and the word rated are in the same language, no systematic effect should be observed. As mentioned earlier (see Method) the compound bilinguals were recalled to the laboratory about 5 weeks after the first testing and the new different-word control condition (same

TABLE 3

AVERAGE CHANGE IN POLARITY OVER THE SUM OF TWO WORDS AND FIVE SCALES FOR THE COMPOUND GROUP ON THREE DIFFERENT CONDITIONS

Condition	Compounds ($N = 28$)[a]			t's Between Conditions	
	Mean	*SD*	t[b]	Same Language	Cross-satiation
Different-word control Different language	−1.29	3.48	1.93*	1.50	2.89***
Different-word control Same language	−0.79	3.14	1.32	—	3.98***
Cross-satiation	−3.14	3.30	4.98***	—	—

[a] Two *Ss* were not available for this part of the testing.
[b] t for departure from zero, two-tailed.
* $P < .10$.
*** $P < .01$.

language) was administered using the same standard procedure as before. The results of this testing as well as those obtained previously with this group are presented in Table 3. It can be seen that both types of different-word control conditions are significantly different from the cross-satiation condition, but they are not significantly different from each other. Although the difference is in the predicted direction (greater negative mean under the different language control), this deduction is not supported to a reliable degree by the present results.

SUMMARY

The phenomenon of semantic satiation was related to the mediation theory of compound and coordinate bilingualism. It was predicted that compound bilinguals, having interdependent language systems, will exhibit a cross-linguistic satiation effect (cross-satiation), while coordinate bilinguals, having functionally independent language systems, would not exhibit the cross-satiation effect. The results supported the prediction that the two types of bilinguals would behave differently and showed that compounds clearly exhibit the cross-satiation effect while coordinates react in the opposite manner. Some additional analyses were discussed which have a further bearing upon the differences in the language structures of compound and coordinate bilinguals.

REFERENCES

CARROLL, J. B., & SAPON, S. M. *Psi-lambda foreign language aptitude battery.* Cambridge, Mass.: Author, 1955.

ERVIN, S. M., & OSGOOD, C. E. Second language learning and bilingualism. In C. E. Osgood & F. Seboek (Eds.), *Psycholinguistics. J. abnorm. soc. Psychol.*, 1954, **49**, Suppl., 139–146.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

LAMBERT, W. E. Measurement of the linguistic dominance of bilinguals. *J. abnorm. soc. Psychol.*, 1955, **50**, 197–200.

LAMBERT, W. E., HAVELKA, J., & CROSBY, C. The influence of language-acquisition contexts on bilingualism. *J. abnorm. soc. Psychol.*, 1958, **56**, 239–244.

LAMBERT, W. E., & JAKOBOVITS, L. A. Verbal satiation and changes in the intensity of meaning. *J. exp. Psychol.*, 1960, **60**, 376–383.

OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

WEINREICH, U. *Languages in contact.* New York: Linguistic Circle of New York, 1953.

(Received September 9, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 583–587

RESPONSE PRETRAINING: DIFFERENTIATION OR AVAILABILITY?[1]

ELI SALTZ

Center for the Study of Cognitive Processes, Wayne State University

Underwood, Runquist, and Schulz (1959) have recently demonstrated that pretraining on the response terms of a paired-associate list facilitated learning of this list. The present paper attempts to examine the mechanisms basic to this effect and to evaluate the relative efficacy of the writer's differentiation theory (Saltz, 1960, 1961) as opposed to the Underwood, Runquist, Schulz position that response pretraining is effective because it increases the availability of the responses.

Underwood, Runquist, and Schulz (1959) have suggested that paired-associate learning is a two-phase process. The first phase involves learning the relevant responses. The second phase involves associating these responses to their appropriate stimuli. Pretraining *S*s on the relevant responses, according to Underwood et al., makes the responses available to *S* and permits the association phase to begin more quickly.

The writer's position is that association is not a critical factor in verbal learning: if *S* had to learn a single pair, learning would occur in very few trials. The critical factor is differentiation of the stimuli and responses to be learned. However, contrary to Gibson's (1940) position, the differentiation focused upon is not between S-R units, but differentiation between stimuli and differentiation between responses. The writer has proposed the following assumptions concerning the differentiation process.

1. Initial differentiation: The initial differentiation of an item, for an *S*, can be defined to be a function of the similarity between this item and other items in a list. The less the similarity, the greater the differentiation.
2. Development of differentiation: Differentiation of the term can be defined as increasing as a function of the frequency with which the term has been presented and responded to (e.g., "familiarization" training, Noble, 1955).
3. Response differentiation: Increased differentiation of the response term, in an S-R pair, is assumed to increase the probability that the term will be emitted as a response—though not necessarily to its appropriate stimulus.
4. Stimulus differentiation: Increased differentiation of the stimulus term is assumed to increase the probability that the stimulus term will elicit its *appropriate* response. In other words, response differentiation is required for emission of a term, but it will not be elicited consistently by its appropriate stimulus until that stimulus is also differentiated.
5. Oscillation of differentiation: Both stimulus and response differentiation oscillate randomly and independently from trial to trial, reducing the effective differentiation from that which would be predicted solely from the initial and developed degrees of differentiation. However, the magnitude of oscillation decreases as a function of increased differentiation.

The specific purpose of the present experiment was to determine if pretraining amounts to anything *more* than making the relevant responses available to *S*.[2] This was done by

---

[1] The present study is part of a program of research, supported by a grant from the National Science Foundation, concerned with the influence of differentiation on verbal processes.

Alfred Klipple ran *S*s for this study, and Shirley Brown assisted in data analysis.

[2] In fairness to the Underwood et al. (1959) position, it should be pointed out that they do not explicitly deny the possibility that factors other than availability may be operat-

making the responses available on cards to all *S*s during paired-associate learning, and forcing *S*s to use these cards during learning (see Procedure section for method used to ensure use of cards). Half the *S*s had received familiarization pretraining on the responses, half had not. If the effect of pretraining is nothing more than making the responses available to *S*s, then there should be no difference between groups receiving pretraining and groups receiving no pretraining. On the other hand, if pretraining is a predifferentiation device, then pretraining should facilitate learning despite the fact that the responses are available to all *S*s.

It also follows, from the writer's formulation, that one of the major effects of response predifferentiation is to reduce oscillation. This was also tested.

METHOD

Materials.—The stimuli and responses used in the present study were the same as those used by Underwood, Runquist, and Schulz (1959), but the pairing of stimuli and responses was not identical with theirs since they did not indicate the specific pairings. The criterion learning materials consisted of two lists of 10 pairs each in which the stimuli were Glaze nonsense syllables from 45–53% association value and of low intralist similarity. The responses were two-syllable adjectives. In one list all responses were relatively synonymous; in the other list the responses were unrelated in meaning.

Subjects.—The *S*s were 80 Wayne State University students from the introductory psychology course and were randomly assigned to compose groups of 20 *S*s in each of the four conditions.

Pretraining.—The *S*s in the pretraining conditions were shown the responses they were to learn in the subsequent paired-associate task, using a procedure similar to that of Underwood et al. (1959). The list of responses was presented in five alternate learning and recall trials. During the learning trials, items were presented at a 2-sec. rate and *S* was instructed not to learn the adjectives in order since they would be in a different order on each trial. Five different orders were used. Following each learning trial, *S* was asked to write down as many responses as he could remember, and was given 1 min. for this.

Availability forcing.—During paired-associate learning, a card with all the relevant responses was placed on the top of the memory drum. The responses were numbered serially from 1 at the top of the card to 10 at the bottom. The *S* was instructed to respond by giving the *number* appropriate to the response he felt was correct. Five different orders of responses were used so that the number associated with any given response changed from trial to trial. This obviated the tendency for *S*s to learn a number-response association, and forced *S* to examine the card for each response. The five orders of items on the cards were different from any of the orders used during pretraining and from any of the orders of responses on the memory drum during paired-associate learning.

General procedure.—Half the *S*s received pretraining on the responses to be learned in their criterion paired-associate task. Since Underwood et al. (1959) obtained very similar results with *S*s who received pretraining on irrelevant responses and those who received no pretraining, in the present study *S*s in the no-pretraining conditions were not given pretraining on irrelevant responses. The paired-associate learning was at a 4:4-sec. rate with an 8-sec. intertrial interval. The 4-sec. rate was necessary since preliminary study showed that *S*s could not examine the availability cards and emit a response in 2 sec. Five different random orders of pairs were used, none of which was the same as the order on any pretraining trial or on any of the availability cards. The *S*s received 15 presentations of the paired-associate list, and anticipated the correct responses on the last 14 of these trials.

RESULTS

As can be seen in Fig. 1, results of pretraining in the present study are very similar to those obtained by Underwood et al. (1959), despite the fact that all *S*s in the present study

---

ing in the response pretraining situation. On the other hand, the writer's position is that response differentiation is a more basic variable from which variables such as availability can be predicted.

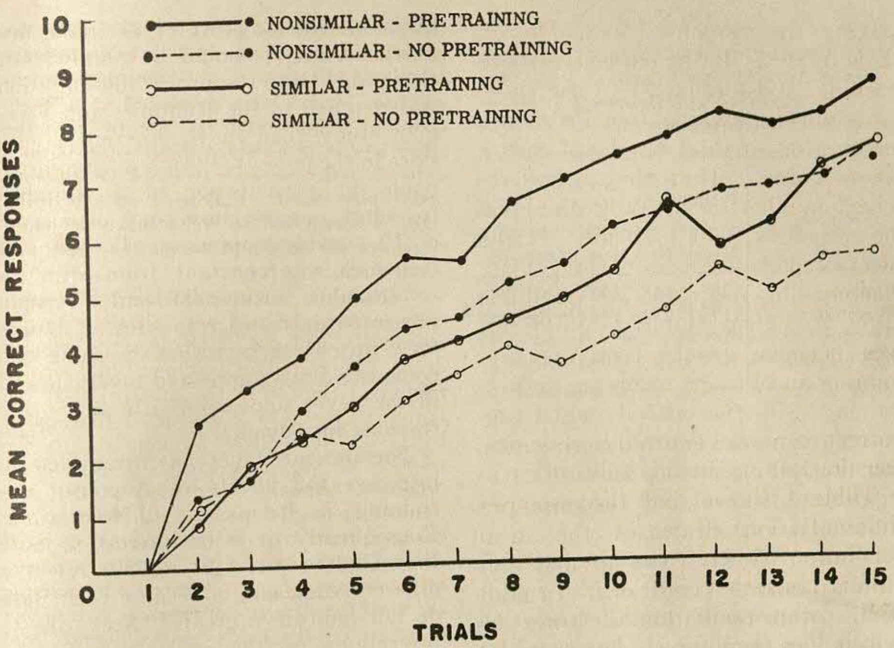


FIG. 1. Mean number of correct responses per trial as a function of response pretraining and response similarity.

had the responses available to them on cards during paired-associate learning, and were forced to use these cards in order to respond correctly. As in the Underwood et al. (1959) study, during the first few trials of paired-associate learning, response pretraining was most effective for *S*s who learned material with low intraresponse similarity; by the end of training, pretraining was facilitating both the high and low response similarity groups. A 2 × 2 analysis of variance over the total number of correct responses for the last half of paired-associate training indicated that the effect of pretraining was significant at beyond the .01 level, the effect of response similarity was significant at beyond the .005 level. The interaction between Pretraining and Response Similarity did not approach significance.

A gross oscillation score was calculated for each *S* by summing the number of responses emitted correctly on each trial, which were *not* correctly emitted on the subsequent trial. This gross oscillation score was then divided by the total number of correct responses given by *S* on all trials prior to the trial on which he made no further errors. This yielded an index of the mean amount of oscillation per correct response. Mean oscillation was computed in this manner because some *S*s (particularly in the nonsimilar pretrained group) reached 100% correct responses early in the 15 trials and made no further errors. Dividing by total number of correct responses for such *S*s would give a spuriously low mean oscillation score which would appear to support the writer's hypothesis, but which would actually be merely a reflection of the great number of

TABLE 1

Mean Oscillation per Correct Response as Affected by Response Pretraining and Response Similarity

Responses	Pretraining		Nonpretraining		Mean of Means
	Mean	*SD*	Mean	*SD*	
Similar	.205	.125	.307	.201	.256
Nonsimilar	.172	.083	.224	.110	.198
Mean of means	.188		.266		

correct responses emitted during overlearning.

Table 1 shows that response predifferentiation decreases the mean oscillation for both similar and nonsimilar response conditions. In addition, mean oscillation is lower for nonsimilar than for similar material. Both these results are consistent with the writer's formulation concerning the effects of differentiation.

A 2 × 2 analysis of variance (Pretraining × Similarity) indicates that the effect of training on mean oscillation is significant at beyond the .02 level. The effect of similarity falls at approximately the .06 level, and so cannot be considered statistically significant. The interaction yielded an *F* of less than 1.00.

DISCUSSION

A recent study by Runquist and Freeman (1960) investigated the effects of response pretraining in a "verbal discrimination" situation: two terms were presented simultaneously, and *S* was to choose one of these. Pretraining on *both* terms in a dyad produced results which were no different from those of a control group with no pretraining. Why did Runquist and Freeman fail to obtain a pretraining effect while such an effect was obtained in the present study? While it is impossible to answer this question with any degree of confidence at this time, it should be pointed out that much less response differentiation was required in the Runquist and Freeman study than in the present study, since in the present study *S* was required, upon the presentation of a stimulus, to select one response out of a pool of 10 possible responses. The pool of responses was constant from stimulus to stimulus. Runquist and Freeman presented only two responses at a time for *S* to choose between; the same two responses always appeared together, and neither ever appeared with any other response in a dyad.

The present paper has attempted to interpret the effects of response pretraining in terms of differentiation. Consequently, it is of interest to note that Underwood et al. (1959) reject a differentiation notion because their data do not indicate a greater effect, due to pretraining, for the similar than for the nonsimilar responses. They reason that, if pretraining reduces similarity by means of differentiating the responses, the effect should be greater for similar than nonsimilar material. While this reasoning may be relevant to Gibson's (1940) notion of differentiation, it is not relevant to the formulation presented in the present paper. Within the present meaning of "differentiation," pretraining is effective largely because of the extent to which it reduces oscillation. The relative effect of pretraining on oscillation at various levels of similarity is an empirical question still to be determined.

SUMMARY

The writer has hypothesized that response pretraining is a predifferentiation device. Underwood (Underwood et al., 1959) has hypothesized that it is a procedure for making responses more available during subsequent association learning. The present study investigated the question, does response pretraining facilitate learning even if the responses are made available to *S* during subsequent association learning. To this end *Ss* were presented with cards containing all the relevant responses, during the association phase of the experiment. Each response was

numbered, and the numbers were randomly reassigned to responses from trial to trial. The *S*s were required to emit the number assigned to the appropriate response. Thus all *S*s were forced to use the availability cards.

The results indicated that pretraining was effective even when the responses were available to all *S*s. One of the major effects of pretraining proved to be the reduction of mean oscillation per correct response.

REFERENCES

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

RUNQUIST, W. N., & FREEMAN, M. Roles of association value and syllable familiarization in verbal discrimination learning. *J. exp. Psychol.*, 1960, **59**, 396–401.

SALTZ, E. Similarity and differentiation in verbal learning. Paper read at Psychonomics Society Meetings, Chicago, 1960.

SALTZ, E. Effect of induced stress on free associations. *J. abnorm. soc. Psychol.*, 1961, **62**, 161–164.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

(Received September 9, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 588–593

THE EFFECT OF STRESS ON STIMULUS DIFFERENTIATION [1]

ELI SALTZ AND WINIFRED J. RIACH

Center for the Study of the Cognitive Processes, Wayne State University

The senior author (Saltz, 1961) has been concerned with the operation of differentiation in human learning. The present experiment is concerned with the effect of stress on such differentiation.

In discrimination learning, degree of differentiation between competing S-R systems is a function of number of reinforcements. Saltz (1961) has postulated that stress acts to break down differentiation between systems. Since the tendency of one system to intrude on another is a function of stimulus similarity (e.g., overlap in common elements between the stimuli to be discriminated), stress should be more disruptive for overlapping systems than for nonoverlapping. This expectation assumes equal differentiation for the overlapping and nonoverlapping systems. Should a greater number of reinforcements be required for differentiation of the overlapping system, the relatively greater disruption of this system resulting from stress would decrease as a function of the greater number of reinforcements. Finally, work by Castaneda and Palermo (1955) suggests that the amount of dedifferentiation produced by stress is greater when original differentiation is low than when it is high.

The following predictions arise from the above formulation. In discrimination learning, when the correct response is dominant, introduction of stress should reduce the probability of correct responses. The greater the similarity between the stimuli to be discriminated, the greater this reduction will be. The reduction will be greatest for similar stimuli when the dominant response is relatively weak.

It should be noted that the predictions arising from the above formulation are different from those made by Castaneda and Palermo (1955), Castaneda (1956), and Palermo (1957), who conceptualize stress in terms of Hullian drive (*D*). Castaneda and Palermo suggest that stress should facilitate performance when the correct response is highest in the hierarchy, since increased stress produces increased drive, and drive will multiply all existing habit tendencies. On the other hand, if an incorrect response is highest in the hierarchy, stress will have detrimental effects. All three cited studies found that in a simple discrimination task, when the correct response was low in the response hierarchy, performance was hindered by stress more than when the correct response was high in the hierarchy. None of the three studies obtained a significant facilitative effect due to stress. It should be noted, however, that they employed increased pacing (e.g., shorter period of time permitted for response emission) to produce stress. It is not certain that this method of producing stress is equivalent to introducing a painful stimulus or to raising the apprehensiveness of *S*.

The present study was designed to investigate the effects of electric shock as stress in discrimination learning.[2]

METHOD

General design.—The *S*s learned a discrimination task to a low criterion of performance, and were then subjected to electric shock simultaneously with the presentation of the discriminanda. A 2 × 2 design was employed: shock-nonshock; similar-dissimilar discriminanda.

---

[1] The present study was supported by a grant to the senior author from the National Science Foundation.

[2] The basic problem in the present study was investigated in a preliminary experiment (Riach, 1958) which showed a decrease in correct responses with shock stress, but was otherwise uninterpretable for reasons of faulty design.

Subjects.—The *S*s were 80 volunteers from the introductory psychology course at Wayne State University. They were randomly assigned to groups of 20 *S*s (10 men, 10 women) in each of the four experimental conditions.

Apparatus.—The apparatus was modeled after one described by Estes and Burke (1955). Stimuli in the discrimination task were 12 6-v., .25-amp. light bulbs, arranged in a circle with a 9-in. radius on a 2-ft. square display panel painted flat black (see Fig. 1). The patterns of lights were programed on tape. A 7 × 9 in. box was placed 18 in. in front of the display panel, and held two push buttons mounted 5 in. apart on top of the box. The *S*'s task was to learn which of these buttons to depress upon presentation of any stimulus pattern. A 6-v. reinforcement light was located 3 in. above each button; these lights were used to indicate, after each trial, which button should have been pressed. The *E* was informed of the button *S* had depressed by the illumination of a corresponding light on a panel facing *E*. This light could not be activated if *S* responded after illumination of a reinforcement light.

Electric shock was introduced by means of two electrodes attached to *S*'s wrist.

Tasks.—A pattern of three lights on the display panel was shown to *S* for 2 sec. and *S* was to depress either the right or left button on the response box. When the panel lights went off, the reinforcement light indicated the correct button on the response box for 1 sec. There was a 4-sec. interval before the next sequence began. This sequence, which constituted a trial, was electronically controlled.

The highly similar overlap task consisted of making a left-button response when the majority of the three lights appeared at Positions 1–6 on the display panel (see Fig. 1) and a right-button response when the majority of the three lights appeared at Positions 7–12. Stimulus overlap resides in the fact that any stimulus could become connected to both responses, e.g., Stimulus 3 would be part of a pattern indicating the left-button response if one or both of the other stimuli were from Positions 1–6; on the other hand, Stimulus 3 would be part of a pattern indicating the right-button response if the remaining two stimuli were from Positions 7–12.

The nonoverlap task involved pairing any stimulus light with only one response. A left-button response was indicated when three lights at Positions 3–8 were activated; a right-button response was called for by lights at Positions 9–2. Different splits of the lights were used for the overlap and nonoverlap conditions because preliminary work indicated that a vertical split (1–6 vs. 7–12) was too easy for the nonoverlap condition, while a diagonal split (3–8 vs. 9–2) was too difficult for the overlap condition.

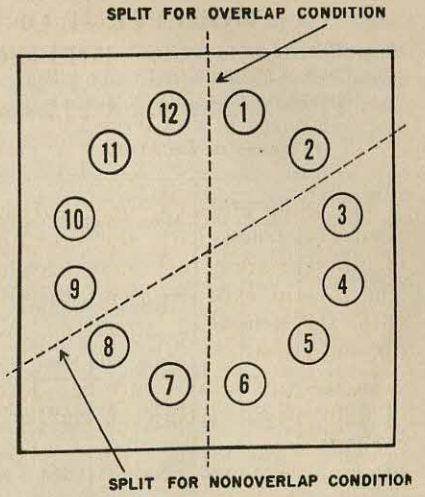


FIG. 1. Illustration of display panel. (Numbers indicating positions of lights, and broken lines indicating splits used for overlap and nonoverlap conditions, are indicated to facilitate understanding of experimental procedure, and were not present on actual display panel.)

Procedure.—The shock device was attached to the nonpreferred hand of each *S*, whether in shock or nonshock condition, before learning was to begin, and shock was adjusted to a point *S* considered to be the pain threshold. The *S* was then told, "now just forget about the armband because I won't apply any stimulation for quite a while." The *S* was also told: "The light will come on over the button you should have pressed for each pattern. This will tell you if you are right or not." All responses were made with the forefinger of the preferred hand. The *S* could rest this hand on the table in front of the response box between responses.

There were 30 different patterns of lights constituting 30 trials for each task before repetition of stimuli began. Responses were tallied in blocks of 12 trials. Nine or more correct responses out of a possible 12 on two

TABLE 1

MEAN BLOCKS OF 12 TRIALS TO CRITERION AND MEAN NUMBER OF CORRECT RESPONSES ON LAST BLOCK BEFORE AND THREE BLOCKS AFTER EXPERIMENTAL TREATMENT

Group	Blocks of Trials to Criterion		Mean Number Correct			
			12 Trials Before		36 Trials After	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Nonoverlap						
Nonstress	3.95	2.21	10.50	.81	32.00	3.54
Stress	3.80	1.54	10.45	.86	30.25	3.69
Overlap						
Nonstress	6.40	3.17	10.20	1.07	29.35	6.95
Stress	7.15	3.66	10.20	1.15	26.75	6.96

separate (though not necessarily consecutive) blocks of trials constituted the criterion for learning. With no pause or instructional indication that a new phase of the testing had begun, *S*s received 36 trials subsequent to reaching the criterion. For an *S* in a stress condition, shock accompanied the first of these 36 trials, plus 11 other randomly selected trials. Shock onset coincided with the onset of the light pattern for a trial, and was removed after $\frac{1}{4}$–$\frac{1}{2}$ sec., before *S* had an opportunity to make a button pushing response.

RESULTS

Table 1 summarizes the number of blocks of trials to reach criterion (including criterional trials) and the performance of the various groups on the 12 trials on which *S*s reached criterion and on the subsequent 36 test trials. As can be seen, the nonoverlap groups reach criterion somewhat earlier than overlap groups. However, there is no significant difference in the strengths of the correct responses for the various groups on the criterion block of 12 trials. Stress produced a relative drop in performance in both overlap and nonoverlap conditions. Also, during the 36 test trials, both overlap groups performed more poorly than the corresponding nonoverlap groups. A difference score was obtained for each *S* between number correct on the criterion trials and on the subsequent 36 trials, and a 2 × 2 analysis of variance was performed over these difference scores. The *F* for stress was significant at beyond the .02 level, the *F* for overlap was significant at beyond the .05 level, and the Stress × Overlap interaction did not approach significance.

Next, the relationship between amount of discrimination learning before the test trials and the performance on the test trials was investigated. It will be recalled that the criterion of learning, before institution of the test series, consisted of *S* making at least 9 correct responses in each of two blocks of 12 trials. A transformed score was calculated for each *S* which was determined primarily by his performance on the last of the two criterion blocks of trials, but which gave some additional weight to performance on the first criterion block. Scores of 9, 10, 11, and 12 correct on the last criterion block of trials were given minimum transformed scores of 1, 5, 9, and 13, respectively. An additional point was added onto the transformed score for every correct response over 9 on the *first* criterion block of trials. Thus, for *S*s who gave 9 correct responses on the last criterion block: 10 correct on the first criterion block was given a transformed score of 2, 11 correct on the first criterion block was given a transformed score of 3, etc. While this transformation was arbitrary, it correlated relatively well with the actual number of correct responses given in the 36 postcriterion test trials. For the nonshock groups, *r* was .40 for the nonoverlap condition, .58 for the overlap condition. For the shock groups, both overlap and nonoverlap conditions produced *r*'s over .70.

For each of the two nonshock groups, a regression was calculated between transformed criterion performance and test performance. For the nonoverlap group, $Y'=.465X+28.04$ where Y' is predicted test score and X is the transformed criterion score. For the overlap group, $Y' = .71X + 24.88$. Figure 2 shows the relationship between transformed criterion scores and scores on the test trials for the two shock conditions, with the regression line based on the corresponding nonshock conditions indicated. As can be seen, in both overlap and nonoverlap conditions *S*s with high scores on the preshock criterion trials show almost no drop in performance during the 36 test trials as compared with the regression line for the nonshock group. It is the *S*s who learned to a low criterion who show the effects of shock. Also, it is apparent that the overlap low criterion *S*s drop much more compared to the regression line for their nonshock counterparts, than the nonoverlap low criterion *S*s.

Next the relationship between shock effects and preshock criterion scores were evaluated. The equations, indicated above, predicting test trial behavior from criterion scores for the nonshock control groups, were used to predict test trial behavior for each *S* in the relevant shock groups. These predicted scores were then subtracted from *S*s' obtained scores. The discrepancy scores obtained in this way indicate the extent to which shock depressed performance for each *S*. Next, all *S*s in each shock group were divided into two groups, those above the median on the preshocked cri-

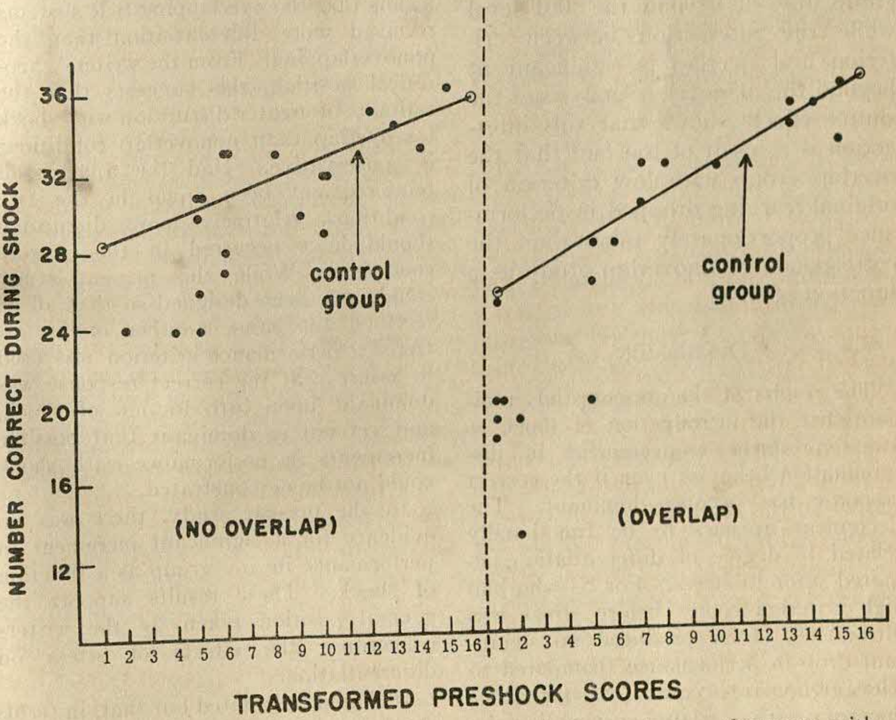


FIG. 2. Scatter plot showing relation between pre- and postshock performance, with regression lines for corresponding nonshock control groups indicated.

terion measure, and those below the median. (It was found that 11 *S*s in each shock group received a transformed score of 6 or lower—transformed score of 6 corresponds to 10 correct on each of the criterion trials—while 9 *S*s in each group received a score of 7 or more, and this split came closest to dividing the two groups in half.) For low criterion *S*s, the mean discrepancy between number correct in the shock condition and number correct predicted on the basis of the nonshock controls was −5.34 in the overlap condition, −2.09 in the nonoverlap condition. For high criterion *S*s, the mean discrepancy was 0.27 for the overlap condition, −0.59 for the nonoverlap. A 2 × 2 analysis of variance indicates that the shock induced drop in correct responses in the low criterion group is significantly greater than that in the high criterion group at well beyond the .001 level while the interaction between criterion and overlap is significant at beyond the .05 level. Analysis of the simple effects shows that this interaction is a result of the fact that the overlap group with low criterion of original learning dropped in performance proportionately more than the corresponding nonoverlap group as a function of shock.

DISCUSSION

The results of the present study indicate that the introduction of shock as stress produces a decrement in discrimination behavior even if the correct response has become dominant. The decrement appears to be functionally related to degree of differentiation acquired prior to stress. For *S*s who had high criterion scores before stress was introduced, stress produced no significant drop in performance (compared to relevant nonstress control *S*s) for either the group with overlapping stimuli or the group with nonoverlapping stimuli. For *S*s with low criterion scores, however, stress produced a significantly greater decrement in the overlap group than in the nonoverlap group, and both groups showed a significant decrement. These latter results must be treated cautiously for two reasons. First, the differences in criterion scores between high and low criterion *S*s are based on individual differences in performance rather than controlled variation of degree of learning across *S*s. Consequently, it is possible that the greater discrepancy in test performance between shock vs. nonshock low criterion *S*s, relative to shock vs. nonshock high criterion *S*s, is due to some factor related to learning rate rather than degree of differentiation. Secondly, the overlap and nonoverlap tasks were not equally difficult; as can be seen in Table 1, the overlap conditions required more trials to criterion than the nonoverlap conditions. Since it is assumed that differentiation develops as a function of number of trials, this means that the overlapping S-R systems received more differentiation than the nonoverlapping. From the writers' theoretical position, this suggests that the estimate of greater disruption with shock for overlap than nonoverlap conditions is conservative. Had the amount of reinforcement been equal in the two conditions, relatively more disruption should have occurred in the overlap condition. While the present study could have been designed so that all *S*s received the same number of pretest trials, a performance criterion was used to assure that the correct response was dominant upon introduction of shock, and yet not so dominant that possible increments in performance with shock could not be demonstrated.

In the present study, there was no evidence for a significant increment in performance in any group as a function of shock. These results support the general position taken by the writers concerning the effects of stress on differentiation.

It should be pointed out that, in terms of the writers' position, the effect of stress on performance can, under appropriate circumstances, be facilitative since the effect will depend on the degree of differentiation of the correct S-R system and competing systems. If the correct system is strongly differentiated and all competing systems are very weak, stress should have only a small detrimental effect on the differentiation of the correct system, but should obliterate the weak competing responses; the resultant would be an increase in performance. Increases in noxious stimulation which have facilitated performance in running studies (e.g., Campbell & Kraeling, 1953) and eyelid conditioning studies (e.g., Passey, 1948) can be cited as demonstrating this type of relationship. These studies have of course, in the past been cited as evidence for the multiplicative drive properties of noxious stimulation, i.e., performance = $f(H \times D)$. It is interesting that in this case both the Hullian multiplicative assumption and the writers' assumption that stress breaks down differentiation (divisional rather than a multiplicative assumption) can predict the same phenomena.

SUMMARY

The present study was designed to investigate the effects of stress on previously acquired differentiations. The *S* learned to criterion either a discrimination between stimuli with overlapping stimulus elements, or one between nonoverlapping stimuli. Subsequent introduction of stress in the form of electric shock had the following effects: (*a*) Shock produced a decrement in performance. (*b*) Analysis of individual differences in number of correct responses on the criterion block of trials indicated that the *S*s who were at high levels of performance prior to the shock showed little decrement. Most of the decrement was contributed by *S*s who had learned the discrimination to a relatively low (though greater than chance) criterion. (*c*) For *S*s who were at a high level of performance prior to being shocked, stress produced no differential effect between overlapping stimulus conditions and nonoverlapping. However, there was a large and significant difference for *S*s who learned to a relatively low criterion, with overlapping groups showing more decrement than nonoverlap. (*d*) No group showed a significant increment as a function of stress. The results were related to theories of differentiation and drive (*D*).

REFERENCES

CAMPBELL, B. A., & KRAELING, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 97–101.

CASTANEDA, A. Effect of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52**, 9–12.

CASTANEDA, A., & PALERMO, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175–179.

ESTES, W. K., & BURKE, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.

PALERMO, D. S. Proactive interference and facilitation as a function of amount of training and stress. *J. exp. Psychol.*, 1957, **53**, 293–296.

PASSEY, G. E. The influence of intensity of unconditional stimulus upon acquisition of a conditional response. *J. exp. Psychol.*, 1948, **38**, 420–428.

RIACH, W. J. The effect of stress on learned discriminations as a function of stimulus similarity. Unpublished master's thesis, Wayne State University, 1958.

SALTZ, E. The effect of induced stress on free associations. *J. abnorm. soc. Psychol.*, 1961, **62**, 161–164.

(Received September 9, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 594–597

FAMILIARITY AND APPARENT COMPLEXITY OF RANDOM SHAPES[1]

ALVIN G. GOLDSTEIN

University of Missouri

This study attempts to show that judgments of visual complexity are affected by stimulus familiarity. Typically, in previous experiments, perceptual complexity has been defined in terms of the physical characteristics of the stimulus (see Hake, 1957, pp. 59–69 for an excellent review of the many definitions and behavioral measures of figural complexity; also see Attneave, 1955; Elliott, 1958; Fitts, Weinstein, Rappaport, Anderson, & Leonard, 1954; French, 1954; Gregg, 1954; Vanderplas & Garvin, 1959). Complexity is a concept which may also be meaningfully specified by variables associated with *S*. This experiment tested the hypothesis that there is an inverse relationship between the apparent complexity of visual shapes and *S*'s familiarity with these shapes.

METHOD

Stimuli.—Ten random shapes were selected from each of the six categories of stimulus complexity used by Vanderplas and Garvin (1959) in their study of verbal associations to random shapes. The six levels of stimulus complexity were 4, 6, 8, 12, 16, and 24 points.[2] All stimuli were black on white 2 × 2 in. slides. A white cardboard screen 60 in. wide by 40 in. high served as the projection surface. A self-timed, automatic projector was used to present the stimuli. The shapes differed in projected size; the longest axis measured from 12 to 18 in.

Learning trials.—The experiment was performed in two sessions. For Session 1, two shapes from each of the six complexity categories were selected using a table of random numbers. A single capital letter (A to M excluding I) was printed in the lower left corner of these slides. This letter was the "name" of the shape. The *Ss* were asked to learn the letter associated with each shape. Trials 1, 2, 4, 6, and 8 were learning trials; *Ss* simply tried to learn the letter-shape pairs. Trials 3, 5, 7, 9, and 10 were test trials; the shapes were presented without the letters and each *S* tried to write the correct letter on a special answer sheet. To increase motivation to learn, prizes of $5, $2, and $1 were given to the *Ss* with the three highest correct identification scores.

Complexity judgments.—During Session 2, which took place 48 hr. after the learning and test trials, *Ss* were shown 36 new random shapes *plus* the 12 shapes (now without identifying letters) they had become familiar with during the previous session. For each of these 48 shapes *Ss* were required to judge complexity on a 7-point scale. These instructions were read to the *Ss*: "You will be shown black shapes on a white background. I want you to judge how complex each shape is when it is shown to you. Please use the categories listed across the top of your rating sheet."

---

[1] This research was performed under contract with the International Business Machines Corp. under AF33 (600)31315 with the Aeronautical Systems Center of Air Materiel Command.

[2] There appears to be a discrepancy between what Attneave (1957) calls complexity and what Vanderplas and Garvin (1959) believe Attneave said. Attneave (1957) analyzed the variable, ". . . actual number of turns . . ." (p. 222) whereas Vanderplas and Garvin (1959) say Attneave found that "Complexity refers to the number of points which determine inflections on the perimeter of the shape" (p. 147). Although both investigators seem to be *saying* the same thing, Vanderplas and Garvin do not *mean* exactly what they say. Photographs of all their experimental shapes, arranged in six nominal complexity categories, were presented in the published report. For many shapes the actual number of points is less than the nominal number (e.g., 8-point shapes with five, six, or seven turns). This occurred, of course, as a result of the way in which such random shapes are constructed (Attneave & Arnoult, 1956). Fortunately, this argument (if it is one) is not crucial to the present experiment.

Inquiries were not answered at this time. Following Attneave's (1957) procedure, which was designed to acquaint *S*s with the range of stimuli to be judged, each of the 48 shapes was shown for 2 sec. before judging began. At this time *S*s' questions were honored, but answers were only in terms of the instructions. Immediately following this procedure, all the shapes were presented (in a new random order) for 7 sec. each during which time *S*s made their complexity judgments. The rating scale was the same one used by Attneave (1957): extremely simple, very simple, simple, medium, complex, very complex, and extremely complex.

At the end of Session 2 *S*s were asked to complete a familiarity scale on each shape. Each shape was now shown for 5 sec., and *S*s judged whether a shape was (*a*) definitely seen before (that is, before this session), or (*b*) never seen before, or (*c*) uncertain. The purpose of the "uncertain" category was to determine whether shapes other than the learned shapes were thought to be familiar because of similarity between shapes. It must be remembered that simple, randomly constructed shapes often do look very much alike.

An interesting sidelight developed as a consequence of using the familiarity scale. During the second session of a pilot study, identical to the present experiment in procedural details, *S*s were presented with 48 shapes all different from the 12 they had learned in the first session. When the responses to the familiarity scale were analyzed, it was found that for many of the shapes a large number of *S*s had checked "definitely seen before." This finding is very similar to Arnoult's (1956) observation that *S*s work on the principle, "When in doubt, call it familiar" (p. 273).

Subjects.—Two experiments are reported here. The data of the first experiment were obtained from 26 men and women college students. Replication of the experiment was performed on an independent group of 29 men and women students. Each group learned a different set of 12 shapes but both groups were presented with the same 36 unfamiliar shapes. For both groups the sequence of familiar-unfamiliar shapes was identical.

RESULTS

Complexity judgments.—The data of both groups were combined and analyzed together. Therefore there were four familiar and six unfamiliar shapes from each of the six complexity categories. If familiarity has no effect on judgments of complexity, then there should be no statistically significant difference between the mean complexity judgments of the familiar shapes and the unfamiliar shapes.

TABLE 1

MEAN COMPLEXITY JUDGMENTS OF FAMILIAR AND UNFAMILIAR RANDOM SHAPES

No. of Points	Mean Complexity Judgments		
	24 Familiar Shapes	36 Unfamiliar Shapes	All
4	0.75	0.97	0.86
6	1.56	1.71	1.64
8	2.30	2.46	2.38
12	3.10	3.90	3.50
16	3.81	4.30	4.06
24	4.70	5.70	5.20
Mean	2.70	3.17	

Note.—The scale of complexity was as follows: extremely simple, 0; very simple, 1; simple, 2; medium, 3; complex, 4; very complex, 5; extremely complex, 6.

In order to test statistically this proposition, arbitrary values of 0 (extremely simple) to 6 (extremely complex) were assigned to the physical complexity categories. For each shape a mean complexity rating was obtained by tallying all responses according to this system and then dividing by the number of responses (see Attneave, 1957, p. 223 for a discussion of the use of a mean rating). The mean ratings for all the 4-point familiar shapes were then summed and a mean of means was obtained. This procedure was carried out for all the remaining categories of familiar and unfamiliar shapes and the means are presented in Table 1.

The data in the last column of the table clearly support Attneave's (1957) results; psychological complexity is directly related to the number of points used in constructing the random shapes. Four-point shapes were on the average judged to be almost but not quite "very simple" whereas the 24-point shapes were judged to be a little more complex than "very complex."

In every case the mean complexity judgments for the familiar shapes is less than for the unfamiliar shapes. The mean difference of .47 between familiar and unfamiliar shapes when tested by a *t* test for correlated scores results in a *P* between .05 and .02 ($t = 3.189$; 5 *df*). Separate analysis of the data of the two groups of *S*s results in essentially identical results.

Familiarity judgments.—The learning trials were specifically designed to allow *S*s to overlearn the shapes, or at least to learn them well enough to be able to discriminate the learned shapes from the new shapes shown during Session 2. Familiarity ratings were used to determine whether this goal was accomplished. Analysis of these data indicated success; 84% of the combined groups' responses to the 36 unfamiliar shapes were "never seen before," 5% were "uncertain," and 11% were "definitely seen before." Turning to the 24 familiar shapes, 10% of the responses were "never seen before," 2% "uncertain," and 88% were "definitely seen before." Thus, there is evidence that the learning trials did indeed make the shapes familiar to *S*s even after 48 hr. had elapsed.

DISCUSSION

It may be concluded that judgments of complexity of random shapes, constructed in a particular manner, are inversely related to the familiarity of these shapes. Although one is tempted to generalize this finding to include, for example, more meaningful figural stimuli, there is little experimental evidence to offer in support of such a position. There is, however, experiential evidence which is common to many individuals. For example, upon first looking at a new, unfamiliar machine, piece of apparatus, or engine, it appears to be a complicated mass of rods, levers, and moving parts all unrelated to each other. It is difficult to understand how it operates. Given only a little knowledge of the instrument, gained perhaps by simply "actively" looking at it for a short time, it seems to become simpler, its parts become figures against a background whereas prior to familiarization there was little figure-ground separation.

It is clear that anecdotal evidence is not what is needed to bridge the generalization gap; research with visual stimuli other than randomly generated figures should be initiated (cf. Gibson, 1951). Although the implications of Attneave's (1957) study are clear, it is a little difficult to believe that the number of turns in the contour of a shape is the major—if not the only—determinant of apparent figural complexity. The present experiment, in addition to demonstrating the importance of stimulus familiarity, also emphasizes the need for more critical testing of both *S* and stimulus variables. In designing experiments to do this, consideration should be given to the classical or single variable type of experiment. If a factorial design is employed, as in Attneave's (1957) study, the judging task may be made too simple. In this particular example, the outstanding visible difference between shapes was the number of turns in the contour and it is reasonable to believe that *S*'s judgments were primarily based on this major perceptible difference. It is suggested here that this "visibility" of an experimental variable could be much reduced in a single factor experiment. Obviously, many small experiments, each involving one variable, would have to be conducted.

Vanderplas and Garvin (1959) and Goldstein (1961) have supported the proposition that there is an inverse relation between judged figural complexity and association value. This relationship may be in part a result of an interaction between complexity and familiarity. Consider the proposition that association value and familiarity are directly related. Thus, a shape should be more familiar if it reminds one of more things or of the same thing more often. If this is assumed to be true, and the present findings suggest that the more familiar shapes appear less complex, then it follows that shapes of high association value should be judged less complex. In other words there is an inverse relation between complexity and association value on a theoretical as well as an experimental basis.

SUMMARY

Judgments of the complexity of 24 familiar and 36 unfamiliar random shapes of matched physical complexity were obtained from 55 *Ss*. Familiarization was accomplished with a 10-trial paired-associate learning session 48 hr. prior to the complexity judging session. The familiar shapes were judged to be significantly less complex than the unfamiliar shapes.

REFERENCES

ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269–276.

ATTNEAVE, F. Symmetry, information, and memory for patterns. *Amer. J. Psychol.*, 1955, **68**, 209–222.

ATTNEAVE, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53**, 221–227.

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.

ELLIOTT, L. L. Reliability of judgments of figural complexity. *J. exp. Psychol.*, 1958, **56**, 335–338.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N., & LEONARD, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1954, **51**, 1–11.

FRENCH, R. S. Identification of dot patterns from memory as a function of complexity. *J. exp. Psychol.*, 1954, **47**, 22–26.

GIBSON, J. J. What is form? *Psychol. Rev.*, 1951, **58**, 403–412.

GOLDSTEIN, A. G. Spatial orientation as a factor in eliciting associative responses to random shapes. *Percept. mot. Skills*, 1961, **12**, 15–25.

GREGG, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

HAKE, H. W. Contributions of psychology to the study of pattern vision. *USAF WADC tech. Rep.*, 1957, 57–621.

VANDERPLAS, J. M., & GARVIN, E. A. The association value of random shapes. *J. exp. Psychol.*, 1959, **57**, 147–154.

(Received September 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 598–604

MEMORY EFFECTS IN CONCEPT LEARNING[1]

EARL B. HUNT

Yale University

In concept learning experiments *S* is shown stimuli selected from a predefined universe. These stimuli (or *instances*) are divided into two or more classes. The task of *S* is to find a rule which can be used to assign any instance in the universe to the appropriate class. When the stimulus universe is defined in terms of possible stimulus variations there is a finite number of possible answers to any concept learning problem (Hovland, 1952). The identification of particular instances as members of specified classes transmits information in the sense that it shows that certain answers cannot be correct. A concept is defined when only one possible answer remains as the definition of each class.

When the instances are presented in series, with only one instance in *S*'s view at any one time, an additional complication is introduced. The information transmitted by the instances must be retained by *S*. From his point of view the concept is not defined until he retains information sufficient to define the concept. This may require presentation of more instances than are necessary to define the concept from *E*'s point of view.

Studies of memory and verbal learning indicate that order of presentation exerts a powerful influence on retention (McGeoch & Irion, 1951). However, there is an important difference between these studies and concept learning. In the usual verbal learning and retention study a particular item (a *key instance*) is inserted in a standard series. After the series is presented there is a test (usually recognition or recall) to determine whether or not *S* can meet some criterion for remembering the key instance. However, in concept learning *S*'s task is to state or use a categorizing rule derived from the information transmitted by the standard series. In order to do this *S* need not remember any particular instance as such. He does have to retain the information transmitted by certain instances.

In order to study information retention, *S*s were presented with a series of instances sufficient to define a concept. The *S*s were then required to identify the class membership of a test series of unlabeled instances, using the categorizing rule they had derived from the first (training) series. The members of the test series were chosen so that their identification depended upon retention of the information transmitted by a key instance in the training series.

EXPERIMENT I

This experiment was designed to provide data revelant to the form of the "information retention curve" as a function of the serial position of the information to be retained.

---

[1] Based upon a dissertation submitted to Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to express his particular thanks to his advisor, Carl I. Hovland, and to the members of the advisors' and readers' committees. The experiments were performed while the author held General Electric Co. and National Science Foundation fellowships. Research funds were provided by a Ford Foundation grant to Carl I. Hovland for the study of concept learning.

Method

Stimuli.—The stimuli were geometric designs of the type shown in Fig. 1. The particular dimensions and values used to provide stimulus variation (Hovland, 1952) were: type of figure (star, fleur-de-lis, "X", and cross), shape of border (circle or square), color of figure (outline or filled in), and orientation of the "side panels" (0–180°, 45–225°, 90–270°, 135–315°). The side panels are the rectangles extending out from the border of the design.

The designs were dittoed one to an 8½ × 11 in. page. The pages and a cover sheet were stapled into a book. Positive instances of the class to be defined were labeled "Alpha" and negative instances "Bravo."

Information content.—Three sets of instances were used. The "standard set" contained two positive and two negative instances. Three of these were sufficient to eliminate all but two possible concept definitions, "outline figures" or "outline figures and circular borders." The other instance in this set, the key instance, eliminated the definition outline figures. The "redundant instances" were six positive and six negative instances which, taken alone or in combination with the standard set exclusive of the key instance, transmitted sufficient information to restrict the correct definition to either of the two alternatives mentioned. The "test set" consisted of 10 instances whose class membership was not stated. All members of the test set had outline figures and square borders. In order to identify them correctly as negative instances *S* had to retain the information transmitted by the key instance; that instances with outline figures and square borders were not members of the Alpha class.

TABLE 1

DESCRIPTION OF STIMULI USED, EXP. I

Order in Set	Figure	Panel[a]	Color	Border	Type
		Standard Set			
1	X	135	O	C	P
2	X	135	F	C	N
3	X	O	O	S	N
4	Star	O	O	C	P
		Test Series			
1	Star	90	O	S	?
2	Star	135	O	S	?
3	Fleur	135	O	S	?
4	X	90	O	S	?
5	Cross	45	O	S	?
6	Cross	90	O	S	?
7	Star	45	O	S	?
8	Fleur	45	O	S	?
9	Fleur	O	O	S	?
10	Cross	135	O	S	?
		Redundant Instances			
1	Fleur	O	F	C	N
2	Cross	45	O	C	P
3	Star	90	O	C	P
4	Fleur	135	F	S	N
5	Star	90	F	S	N
6	X	O	O	C	P
7	Cross	90	O	C	P
8	Cross	45	F	S	N
9	Fleur	135	O	C	P
10	Cross	90	F	C	N
11	Star	45	F	C	N
12	Fleur	45	O	C	P

Note.—Abbreviations are: O = outline, F = filled, P = positive, N = negative, C = circle, S = square.
[a] The number in this column is the lowest orientation, in degrees, of the dimension "side panel."

Table 1 describes all instances used. The instances are listed in order of presentation within a set. The standard set and the test set were always presented as units. The set of redundant instances was split into two or three blocks as required by the experimental design.

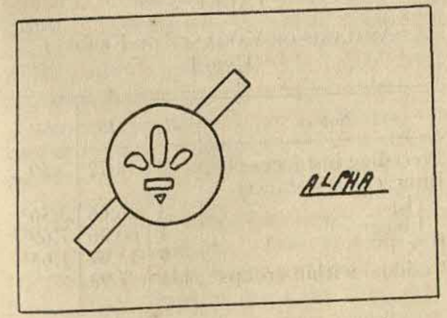


FIG. 1. Typical stimulus: Exp. I.

Design.—The number of instances preceding the key instance and intervening between it and the test series was manipulated by inserting the standard set and the test set into the set of redundant instances. For example, if the standard set is inserted between Redundant Instances 2 and 3 and the test set between Redundant Instances 6 and 7, then four instances are placed in front of the key instance and five intervene between it and the first test instance.

Either 2, 4, 6, or 8 instances preceded the

TABLE 2

Mean Error Scores: Exp. I

Preceding Instances	Intervening Instances				Mean
	1	3	5	7	
2	2.5	3.7	3.6	5.5	3.8
4	3.5	1.8	3.9	4.4	3.4
6	2.2	5.5	4.9	3.8	4.1
8	2.6	4.7	5.6	3.7	4.2
Mean	2.7	3.9	4.5	4.4	

key instance, and 1, 3, 5, or 7 instances intervened between it and the beginning of the test series. These conditions of the two independent variables were combined in a 4 × 4 factorial design.

Subjects.[2]—The *S*s were 160 men and women (10 *S*s per condition) enrolled in the introductory psychology class of Quinnipiac College. The experiment was conducted during regular class periods. Assignment was random within class sections. There were approximately 35 *S*s in each section.

Procedure.—At the beginning of the class period *E* described a concept learning problem, using numbers as stimuli. The *S*s were told that they were to solve a conjunctive concept learning problem (Bruner, Goodnow, & Austin, 1956) which involved no more than two dimensions.

Instances were presented by having *S*s turn the pages of their books on a signal from *E*, which occurred every 30 sec. When *S* encountered a test instance (which was numbered instead of being labeled Alpha or Bravo) he was to indicate, on a separate paper, the class to which the instance belonged.

Redundant instances which were not presented prior to the test series were inserted after it. This made all test books the same length. Since *S* did not know that there would be no further testing after he completed the test series, it was possible, by having all the books the same length, to conduct all experimental conditions simultaneously.

Scoring.—The only conjunctive concept which would correctly categorize all identified instances was "Alpha designs have outline figures and circular borders." If the key instance is excluded, a simpler rule, "All stimuli with outline figures are Alpha," can be derived from the identified instances. It was assumed that in the absence of information to the contrary *S* would utilize the latter, simpler rule. The number of test instances which *S* incorrectly identified as Alpha was used as an operational measure of the difficulty *S* had in retrieving the information transmitted by the key instance. The training series was highly redundant in eliminating possible concept definitions involving other dimensions and values.

This procedure provides a graduated measure of concept learning. It can be argued that such a measure is logically invalid; either *S* has or has not acquired the concept. However, *S*'s (implicit or explicit) statement of a categorizing rule may be only one of several factors in determining his identifications during the test series. Use of a categorizing rule might be masked by "noise" from extraneous variables (e.g., a tendency to match frequency of positive instances in the training and test series). A measure of central tendency would provide maximum sensitivity in determining the information available to *S*. Therefore, the mean number of incorrect identifications was used as the dependent variable.

Results

The mean error scores for the 16 groups are presented in Table 2, Table 3 summarizes an analysis of variance of error scores. Number of errors was a positive linear function of the number of intervening instances. There is no evidence for any effect upon error scores associated with change in the number of preceding instances.

TABLE 3

Analysis of Variance of Errors: Exp. I

Source	*df*	*MS*	*F*
Preceding instances (P)	3	4.72	.59
Intervening instances (L)	3	26.66	3.36*
Linear	1	60.06	7.69**
P × I	9	13.07	1.65
Residual within groups	144	7.93	

* $P < .05$.
** $P < .01$.

---

[2] The *S*s were obtained through the assistance of Robert Evans of Quinnipiac College and Paul A. Lane of the University of Bridgeport.

TABLE 4

MEAN ERROR SCORES: EXP. II

Preceding Instances	Intervening Instances			Mean
	1	9	17	
2	4.1	6.1	7.3	5.8
6	3.2	4.7	4.4	4.1
Mean	3.6	5.4	5.8	

An analysis of types of errors made on the test series did not reveal any striking patterns. In particular, there is no tendency for a particular error to be made in any one experimental condition either more or less frequently than would be expected on the basis of the mean number of errors in that condition.

EXPERIMENT II

In this experiment the number of intervening instances was increased in order to obtain additional empirical knowledge of the form of the information retention curve.

Method

Stimuli.—The stimuli were similar to those used in Exp. I. The dimension "orientation of side panel" was changed to "direction of arrow" and an arrow replaced the "side panel." Triangles, diamonds, and double triangles were added and the fleur-de-lis dropped in the figure dimension. These changes increased the stimulus universe and, hence, the number of redundant instances which could be presented. The stimuli were dittoed and placed on $5\frac{1}{2} \times 8\frac{1}{2}$ in. pages.

Design.—The same type of design was used as in Exp. I. There were 2 or 6 preceding and 1, 9, or 17 intervening instances.

Subjects.—The *S*s were 54 Bridgeport University introductory psychology students (9 *S*s per condition). One protocol was rejected because of improper assembly of a book. This *S* was assigned the mean error score of his group and the error term degrees of freedom reduced by one.

Results

The mean error scores for each condition are presented in Table 4. The corresponding analysis of variance is summarized in Table 5. The only effect to approach conventional levels of significance ($P < .06$) is the positive linear increase in errors with increase in the number of intervening instances. This is consistent with the results of Exp. I. Paradoxically, the number of errors decreases with an increase in the number of preceding instances. However, this effect is not quite statistically significant, and the difference was in the opposite direction for the same number of preceding instances in Exp. I.

EXPERIMENT III

In Exp. I and II differences in serial position of the key instance were confounded with differences in the length of the training series. Also, differences in the availability of information transmitted by the key instance were ascribed to difficulty in retaining information. Before *S* can retain information, he must acquire it. If the acquisition process varies during learning, serial position effects would appear which were not due to information retention.

In particular, the key instance transmitted information about a particular dimension. Perhaps not all dimensions are scanned on all in-

TABLE 5

ANALYSIS OF VARIANCE OF ERRORS: EXP. II

Source	*df*	*MS*	*F*
Preceding instances (P)	1	39.53	3.51
Intervening instances (I)	2	24.80	2.20
Linear	1	44.60	3.95
P × I	2	4.92	
Residual within groups	47	11.26	

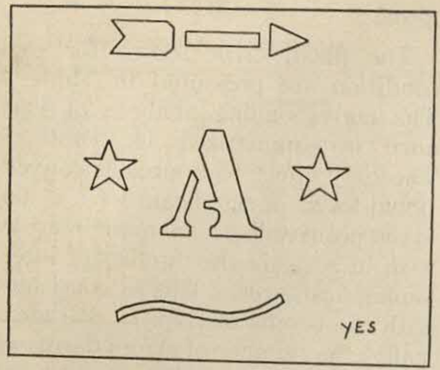


FIG. 2. Typical stimulus: Exp. III.

stances. There is evidence that humans exhibit a systematic order of dominance in scanning dimensions (Dattman & Israel, 1951; Grant, 1951; Heidbreder, 1945). The design of the stimuli and length of exposure interval in Exp. I and II were chosen so that *S*s should have had little trouble in scanning all dimensions. Nevertheless, a more direct test of the effect of transmitting information over different dimensions was desirable.

In Exp. III three key instances, which transmitted information over different dimensions, were presented at various positions in a constant length training series.

Method

Stimuli.—More complex stimuli were used. A typical instance is shown in Fig. 2. There were six dimensions; type of letter in the center (A or C), number of figures on the sides (one or two), type of figure (star, square, or cross), direction of arrow (right or left), number of lines below the letter (one, two, or three), and color of lines (black or white). Training instances were mimeographed on 5 × 8 in. cards, test instances were mimeographed, two to an $8\frac{1}{2} \times 11$ in. page.

Information content.—The training series consisted of 10 positive and 3 negative instances. The negative instances were key instances. Positive instances were marked "Yes" in the lower right hand corner, negative instances were marked "No."

The positive instances all had an arrow pointing to the right, a single figure on either side of the letter, and outline (white) lines below the letter. Thus the positive instances alone restricted the correct concept definition to the inclusive disjunction of these three one-dimensional answers. The three key instances each contained one of the three possible one-dimensional answers, demonstrating that a correct answer involved the union of two or more of the dimensions and values indicated.

The first nine positive instances all had an A in the center. The three negative (key) instances had a C in the center. The tenth positive instance (the thirteenth and last instance of the training series) also had a C in the center. From the results of pretesting, it was felt that *S*s would guess, early in the problem, that "A in the center" was the correct answer. The final instance would be a "positive infirming instance" (Bruner et al., 1956). The *S*s would have to construct a new answer from their memory of previous instances. Comments made by *S*s indicated that the final training instance did force them to develop a new answer. The *S*s were also asked to write a verbal statement of the concept they had derived from the training series. Only 2 *S*s (in different groups) gave answers indicating that they did not realize that the final training instance eliminated the answer "A in the center." These *S*s were replaced.

The test series contained 18 instances. Six of these had an arrow pointing to the right but did not have a single figure beside the letter or outline lines below it. Thus these instances were "positive" by a concept which could be derived from the training series exclusive of one of the three key instances. Similar sets of six test instances were included for the other two key instances.

Design.—Identity and position of key instances (key instances were placed in Serial Positions 1, 5, and 9 of the training series) were combined in a 3 × 3 latin square. In one sequence of training instances Key Instance A appeared in Position 1, B in Position 5, and C in Position 9. This sequence is represented in the first row of the mean error score table (Table 6). The other two sequences of key instances (represented by the second and third rows of Table 6) were CAB, BCA. Each training sequence was given to 10 *S*s, in an example of Lindquist's (1953) Type II design.

Subjects.—Thirty Quinnipiac College students served as paid *S*s. The *S*s were tested in groups of 3 to 8.

Procedure.—Each *S* was given a deck of training stimuli and a test booklet. The same instructions were given as in Exp. I and II except that there was no practice problem. The *S*s from each of the three groups served simultaneously.

Results

The dependent variable was the number of erroneous identifications in the test series which were inconsistent with a particular key instance. The mean number of errors of each type is presented in Table 6, the accompanying analysis of variance in Table 7. The results are not inconsistent with the previously obtained data. Serial position had an effect with a constant length training series. Unfortunately, a stronger statement is not possible because the latin square assumption of noninteracting effects may or may not have been met (Gourlay, 1955). This assumption can be tested by the between-groups and square-uniqueness terms. While neither reaches conventional levels of significance, both are larger than would be desired.

An interaction effect could be produced by selective order in the examination of dimensions. The complex stimuli used should favor some sort of orderly search pattern. This interpretation is consistent with the significant effect of instances, which reflects a reliable difference in the ease with which information is assimilated from different dimensions.

TABLE 6

MEAN ERRORS INCONSISTENT WITH PARTICULAR KEY INSTANCES: EXP. III

Sequence	Position of Instance			Mean
	1	5	9	
ABC	5.0	.5	.2	1.9
CAB	1.8	1.2	.0	1.0
BCA	1.2	2.1	2.2	1.8
Mean	2.7	1.3	.8	

TABLE 7

ANALYSIS OF VARIANCE: EXP. III

Source	*df*	*MS*	*F*
Between *S*s			
Sequence (Groups)	2	7.55	2.46
Within groups	27	3.07	
Within *S*s			
Serial position	2	28.32	5.23*
Instance	2	38.41	7.09*
Sq. uniqueness	2	17.01	3.14
Residual	54	5.42	

* $P < .01$.

DISCUSSION

Considered together, the three experiments present a consistent picture of memory effects during concept learning. The number of errors in identification of instances following presentation of an information-bearing instance increases linearly with the number of instances intervening between the presentation and the test series. The number of instances preceding the key instance does not appear to have a strong effect upon *S*'s ability to retain transmitted information.

These findings may be compared to a related study of memory in concept learning by Cahill and Hovland (1960). They claim evidence for primacy and recency in information retention. Their recency evidence is comparable to "effect of intervening instances." As primacy evidence they show a change in the slope of the function relating probability of use of information to number of intervening instances as the number of preceding instances is varied. This would appear as an interaction effect in the present study. Such an effect was not demonstrated.

Cahill and Hovland (1960) used, as a dependent variable, *S*'s verbal statement of a categorizing rule. In their training series they used only negative instances. In the studies reported here, actual classification behavior was observed and both positive and negative training instances were used. To the extent that the two studies have findings in common, their generality is increased. Further research is required to determine the reason for the difference in the findings on the effect (or lack of effect) of preceding instances upon information retention.

SUMMARY

The utilization of information transmitted by a particular instance was investigated as a function of the position of the instance in a series of instances used to define a concept. Utilization of information was determined by observing *S*'s action on a categorization task. An interference effect was found for instances intervening between the information transmitting instance and the beginning of the test.

REFERENCES

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

CAHILL, H., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

DATTMAN, P. E., & ISRAEL, H. E. The order of dominance among conceptual capacities: An experimental test of Heidbreder's hypothesis. *J. Psychol.*, 1951, **31**, 147–160.

GOURLAY, N. *F*-test bias for experimental designs of the latin square type. *Psychometrika*, 1955, **20**, 273–288.

GRANT, D. A. Perceptual versus analytic responses to the number concept of a Weigl-type card sorting test. *J. exp. Psychol.*, 1951, **41**, 23–29.

HEIDBREDER, E. Toward a dynamic theory of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **59**, 461–472.

LINDQUIST, E. F. *Design of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1951.

(Received September 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 605–611

POSTREST MOTOR LEARNING PERFORMANCE AS A FUNCTION OF DEGREE OF LEARNING[1]

JOHN C. JAHNKE

Miami University

There are extant a number of studies on the effect of amount of prerest practice (degree of learning) on motor learning performance after a rest period (e.g., Adams & Reynolds, 1954; Ammons, 1947; Bell, 1942; Irion, 1949; Kimble, 1949; Wasserman, 1951). However, almost all such studies have involved what might be called massed practice (MP) conditions. The most extreme MP condition in the studies cited above was continuous practice. Less extreme MP conditions involved either 25-5 or 15-5 sec. work-rest cycles; the most highly spaced conditions involved equal amounts of work and rest—either a 60–60 or a 30–30 sec. work-rest cycle. One study (Jahnke, 1958) has explored the effect of degree of learning on postrest performance under practice conditions involving smaller amounts of work than rest with a 15–45 sec. work-rest cycle. Such conditions, involving smaller amounts of work than rest, may be designated by the generic term, distributed practice (DP) conditions. Another DP (15–30 sec. cycle) study by Buxton (1943) also bears on the effect of the degree of learning variable, although the *S*s were practiced to a criterion rather than given a particular amount of practice.

Briefly, the findings of these two studies suggest that the effect of degree of learning under DP conditions is markedly different from its effect under MP conditions. Because of the importance of this suggestion, the two experiments reported below were performed.

METHOD

Experiment I.—Four groups of 20 undergraduate men each were practiced on a pursuit rotor. The naive *S*s were drawn largely from introductory psychology classes. Rate of rotation of the rotor was 60 rpm. Its direction of rotation was clockwise, and it was equipped with a hinged, brass stylus. The brass target was 1 in. in diameter, its surface was flush with the turntable surface, and its center was located 3¾ in. from the center of the 11½-in. diameter turntable. Each group received either 8, 17, 22, or 30 prerest trials with the 10–20 sec. work-rest cycle. All *S*s then received 48 hr. rest and 21 additional 10–20 trials. The long rest was introduced to minimize the complicating effects of any residual, undissipated work inhibition, such as Hull's (1943) reactive inhibition (I_R), on postrest performance. The four groups were matched *S* for *S* on the basis of total time on target scores on the first five trials. A running total was kept for each *S* during the matching trials, and *S* was assigned to one of the experimental groups during the 20-sec. rest following the last matching trial. Time on target scores were recorded to the nearest .01 sec. and were read from a Standard Electric clock. Work and rest periods were cycled automatically by a Lafayette Repeat Cycle timer, Model No. 1. The instructions given were those of a previous study (Jahnke & Duncan, 1956).

Experiment II.—Since the results of the lowest degree of learning in Exp. I were the most striking, greater attention was paid in Exp. II to the lower part of the degree of

---

[1] This research was facilitated by a grant from the Faculty Research Committee of Miami University.

The writer is indebted to S. A. Switzer, who critically read this report. The writer also wishes to acknowledge the assistance of Clifford Hammer, James Korn, and Douglas Sheeran with the collection of data for Exp. I and the assistance of Douglas Sheeran again with the collection of data for Exp. II.

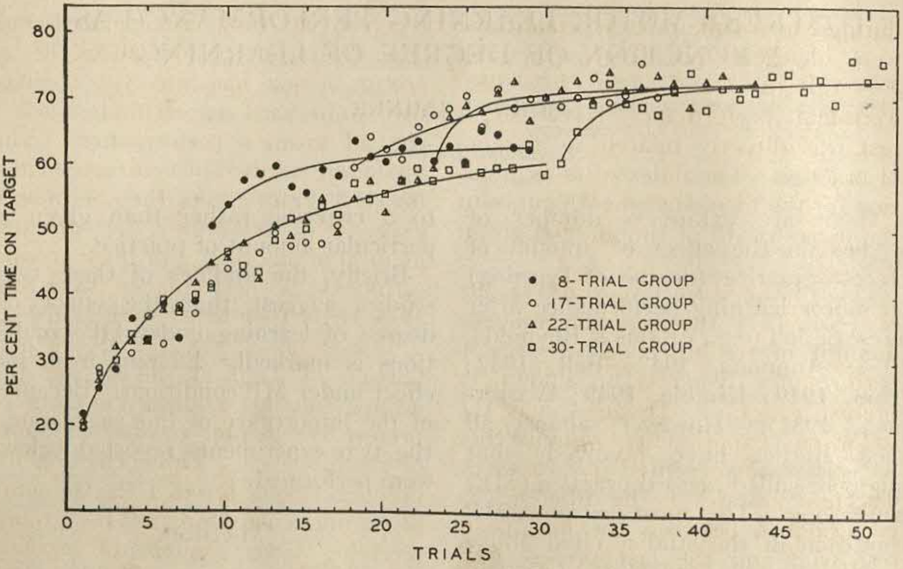


FIG. 1. Performance data for all groups: Exp. I.

learning dimension. Also, the entire dimension was broadened somewhat. Five groups of 20 undergraduate men each were matched *S* for *S* on the basis of total time on target during the first four trials. Again a running total was kept for each *S* during the matching trials, and *S* was assigned to one of the experimental groups during the 20-sec. rest following the last matching trial. All groups were given a total of 60 trials on the pursuit rotor. Each group received either 4, 8, 16, 24, or 36 practice trials, 48 hr. rest, and its remaining trials. Practice was always with the 10–20 sec. work-rest cycle.

The pursuit rotor was modified somewhat for Exp. II. Its target was $\frac{11}{16}$ in. in diameter and was set flush with the turntable surface. The center of the target was located $3\frac{1}{8}$ in. from the center of the $10\frac{1}{4}$-in. diameter turntable. A spring-loaded stylus with a ball-bearing tip was used, but other details of procedure were the same as for Exp. I.

RESULTS

Experiment I.—The performance data and the smoothed curves fitted to portions of the data are presented in Fig. 1. Inspection of the data shows that the performance curves are quite regular, with a few minor exceptions. The smooth curves fitted to the prerest data of the 30-trial group and the postrest data of all groups are from a single family: a hyperbola with an added constant. (This function was also fitted to the data of Exp. II.) The general equation for this family of curves is:

$$Y = \frac{X - 1}{k + mX} + c,$$

where Y = percentage time on target, X = ordinal number of the practice trial, c = performance on the first trial for which the curve is plotted, and m and k are empirically-determined constants. The curves were fitted by the procedure given by Lewis (1960), and both inspection and tests of goodness of fit suggest the fits are reasonably good.

Inspection of the initial portion of the data of Fig. 1 shows that the matching procedure adequately equated the initial performance of the four groups. A related-measures analysis of variance of performance of these groups over the first eight trials, during which time practice conditions were identical for all groups, gave $F = 1.06$ for 3 and 57 *df* ($P > .05$). Terminal performance level before rest was directly related to amount of practice. The index of determination for the fit of the smooth curve to the prerest data of the 30-trial group was .95.

One of the more important results of Exp. I is the function relating amount of reminiscence, or gain in performance over rest, to amount of prerest practice. (In this and the following experiment, reminiscence measures are presented in two ways. The first measure is the difference in performance between the first postrest trial and the last prerest trial for each group. This is a measure of reminiscence uncorrected for practice effects occurring on the first postrest trial. The second measure corrects for practice effects and is the performance difference between the first postrest trial for each group and the trial of the same ordinal number for a control group which has as yet received no long interpolated rest. The averaged performances of all groups which had not yet received rest were used to estimate such a control group's performance. This procedure resulted in a corrected value for each group except the one in each experiment receiving the greatest amount of practice. For these groups, estimates of performance were obtained from their fitted prerest curves.)

Inspection of Fig. 1 and the first portion of Table 1 shows that the groups having smallest amounts of prerest DP exhibit greatest amounts of reminiscence. The correction for practice effects lowered the amounts of reminiscence more for the groups receiving lower degrees of learning than for those receiving higher degrees. All groups of Exp. I show gain in performance over rest when the reminiscence measure uncorrected for practice effects is used; all but the 30-trial group show gain in performance over rest when the alternative reminiscence measure is used. The direct difference *t* values computed sepa-

TABLE 1

SELECTED PERCENTAGE TIME-ON-TARGET MEASURES FOR ALL GROUPS

Group	Last Prerest Trial		First Postrest Trial		"Estimated" First Postrest Trial[a]	Warm-up[b]
	Mean	*SD*	Mean	*SD*	Mean	Mean
Exp. I						
8-trial	38.95	11.79	50.68	15.30	42.52	9.12
17-trial	51.29	15.02	59.64	12.69	55.85	6.28
22-trial	57.64	15.91	59.34	16.41	57.66	11.62
30-trial	58.80	12.12	60.34	16.03	60.49	9.09
Exp. II						
4-trial	17.92	8.89	27.38	11.07	18.88	8.20
8-trial	22.28	6.96	29.30	12.39	22.17	6.44
16-trial	26.14	8.98	30.46	8.93	33.03	12.70
24-trial	36.23	11.48	34.49	11.73	36.77	13.33
36-trial	43.31	12.87	35.90	9.25	40.46	11.91

[a] Estimates of performance on Postrest Trial 1 if no rest had been given. The difference between this score and the score on Postrest Trial 1 for each group yields a reminiscence measure corrected for practice effects. See text for further details.

[b] These values yield performance on Postrest Trial 5 for each group when added to scores obtained on Postrest Trial 1.

rately for each of the four groups, in order of increasing amounts of practice, for the reminiscence measure uncorrected for practice effects were 5.662, 2.673, .464, and .416, with 19 *df* in each case. The values for the 8- and 17-trial groups are highly significant; the remaining values fail of significance.

Warm-up, defined as improvement in performance over the first five postrest trials, was not found to be systematically related to amount of prerest practice for the conditions of this study. It can be seen in Fig. 1 and again in the warm-up data of the Exp. I section of Table 1 that the initial slope of the curve for the 8-trial group is greater than that for the 17-trial group, lesser than that for the 22-trial group, and about the same as for the 30-trial group. However, when warm-up measures were corrected for practice effects occurring during these five trials, greater ordering was obtained. (Corrections for practice effects for this and the following experiment were estimated from the smoothed prerest curve of the group receiving the greatest amount of prerest practice. The improvement in performance over the appropriate stage of practice for this group was subtracted from the uncorrected warm-up score for each group separately.) The warm-up scores corrected for practice effects were 3.82%, 3.88%, 9.91%, and 7.98% time on target for the groups, given in order of increasing amounts of prerest practice. With the exception of the 30-trial group, whose warm-up measure was somewhat less than that of the 22-trial group, warm-up was directly related to amount of practice. Additionally, it is noted that the groups with the two lowest degrees of prerest learning showed similar amounts of warm-up and that the remaining groups showed larger but also similar, amounts.

Asymptotic performance values were obtained for all four groups. It can be seen in Fig. 1 that the 17-, 22-, and 30-trial groups seem to be converging toward the same value. Asymptotic values for these groups determined from the equations for the fitted curves were 78.44%, 72.75%, and 74.37% time on target, respectively. The comparable computed asymptote of the 8-trial group is much lower (65.11%). The fits of the curves to the postrest data seem acceptable in general, although the fit for the 17-trial group is poor. Indices of determination for these curves were as follows: .80 for the 8-trial group, .29 for the 17-trial group (this low value seems attributable largely to poor performance on postrest Trials 4 and 9), .91 for the 22-trial group, and .82 for the 30-trial group. Additionally, a related-measures analysis of variance was performed on the last postrest trial data. These differences were highly significant ($F = 6.62$, 3/57 *df*, $P < .01$).

Experiment II.—The performance data and smoothed curves fit to portions of the data are shown in Fig. 2. The curve-fitting procedure was identical to that of Exp. I. All curves are again hyperbolic in form. Inspection of the initial portion of the data shows that the matching procedure equated the early performance of the five groups. A related-measures analysis of variance of matching trials performance gave an $F < 1$. Terminal performance level before rest was directly related to amount of practice. The index of determination was found to be .89 for the fit of the smoothed curve to the data of the longest prerest practice group.

The findings bearing on reminiscence in Exp. I were essentially re-

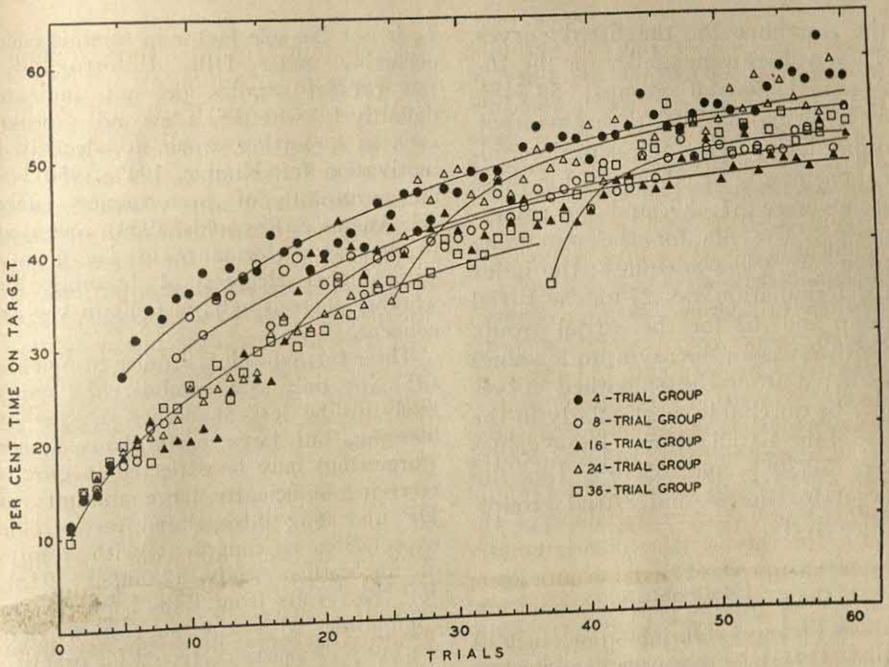


FIG. 2. Performance data for all groups: Exp. II.

peated in Exp. II. Inspection of Fig. 2 and the second portion of Table 1 shows that again groups receiving the smallest amounts of prerest DP showed greatest amounts of reminiscence. In Exp. II, however, losses over rest (forgetting) occurred in the two groups receiving the largest amounts of practice. Direct difference *t* values were obtained separately for the reminiscence gains (uncorrected for practice effects) for each of the five groups. These values were 4.441, 2.686, 2.227, −.753, and −2.714, with 19 *df* in each case, for the groups in order of increasing practice. All but the fourth value (24-trial group) were significant at or beyond the .05 level.

Warm-up scores, defined as for Exp. I as the gain in performance over the first five postrest trials, are shown in the second portion of Table 1. Warm-up was found to be greatest in the groups with the three highest degrees of prerest learning. As for Exp. I, warm-up values were also corrected for practice effects occurring over the first five postrest trials. The corrected values were 2.46%, 2.18%, 10.08%, 11.56%, and 10.81% time on target for the groups, given in order of increasing amounts of prerest practice. Again the two groups with the lowest degrees of learning showed similar amounts of warm-up and the remaining groups showed larger, but also similar, amounts.

The later postrest practice performances of the five groups do not appear to be systematically ordered. A related-measures analysis of variance of total performance over the last 10 trials showed that performance differences at this stage were not significant ($F=2.06$, 4/76 *df*, $P>.05$). Asymptotic performance values for each group were also determined from the equations for the fitted curves. These values were similar for the 16-, 24-, and 36-trial groups: 53.34%, 59.12%, and 53.52% time on target, respectively. The indices of determination for the fitted curves for these groups were .87, .89, and .96, respectively. The fits for the remaining groups were less adequate; the index of determination was .27 for the 4-trial group and .62 for the 8-trial group. For this reason, the asymptotic values predicted from the smoothed curves may be unreliable, particularly in the case of the 4-trial group. These values were 70.30% and 65.40% time on target for the 4- and 8-trial groups, respectively.

DISCUSSION

A temporary work inhibition, such as Hull's (1943) I_R, has been conceived to build up rapidly with repeated responding and to dissipate spontaneously with rest. This hypothesis has been quite useful in dealing with important motor learning phenomena, notably the relative depression of performance under MP and, as relates to the present studies, gains in performance occurring over rest. There has been little question of the utility and validity of the concept of I_R; however, results are accumulating which pose difficulty for the concept.

Studies using highly spaced practice conditions apparently do not produce the direct relation between reminiscence and amount of prerest practice demanded by the I_R principle. In both present studies and in an earlier study (Jahnke, 1958), an inverse monotonic relationship between reminiscence and amount of prerest DP was found. Consistent with this finding is that of Buxton (1943), who demonstrated, depending upon response measure, either an inverse monotonic or an inverted U shaped relation between reminiscence and level of mastery. The results of these DP studies do not of themselves negate the I_R principle. They do, however, show that I_R is not the sole factor in reminiscence occurring under DP. Unfortunately, the present studies do not indicate definitively whether a second process, such as forgetting, warm-up, decreased motivation (cf. Kimble, 1949, 1950), or an asymptote of performance effect (cf. Adams & Reynolds, 1954), operated to depress performance in the groups with higher degrees of learning, or whether there is a basic fault in the I_R concept.

Three further points require comment: (*a*) Not only are reminiscence gains likely to be less at higher degrees of learning, but large performance losses (forgetting) may be expected in groups receiving sufficiently large amounts of DP and long interpolated rest. This expectation is congruent with results of an earlier study (Jahnke, 1958). (*b*) The results from Exp. I and II are compatible with the suggestion that warm-up measures corrected for practice effects are directly related to degree of learning. The data corrected for practice effects all suggest that the transition from relatively small to relatively large amounts of warm-up is abrupt, occurring between Trials 17–22 in Exp. I and Trials 8–16 in Exp. II. (*c*) The final point concerns the estimation of asymptotic values from the curves fitted to the postrest trials of the present study. It will be recalled that the asymptotic behavior of the groups of Exp. I suggested that interruption of practice differentially depressed the performance of the group with the lowest degree of prerest learning, while Exp. II showed that this was not necessarily the case. In order to check on the possibility that the small number of postrest trials of Exp. I might have a marked influence on predictions of the asymptotes, the postrest data of Exp. II were fitted as for Exp. I from the first 21 trials. While these curves are not presented here, the results confirm those of Exp. I. It was found that the asymptotes for the groups with the three lowest degrees of learning were very similar (46.33%, 43.63%, and 44.19% time on target for the 4-, 8-, and 16-trial groups, respectively) and lower than the asymptotes for the two groups with the higher degrees of learning (52.48% and 52.97% time on target for the 24- and 36-trial groups, respectively). Thus, the asymptotic effects noted in Exp. I appear to be attributable to the relatively small number of postrest trials to which the curves were fitted.

While the results of the experiments of this study are not entirely clear on this point, it would seem a reasonable conclusion that groups interrupted at different stages of practice ultimately reach similar postrest performance levels if sufficient postrest practice is given.

SUMMARY

Two experiments were performed to study the effect of degree of learning on postrest performance on the pursuit rotor. The major results were: (*a*) Both prerest and postrest performance data were fitted by a single family of curves. (*b*) Reminiscence was found to be an inverse monotonic function of amount of prerest distributed practice. The implication of this finding for a concept of temporary work inhibition which accumulates with practice is discussed. (*c*) While little or no performance loss over rest (forgetting) was observed in Exp. I, losses were observed in Exp. II. Forgetting occurred in the two groups given large amounts of practice, but was statistically significant only in the group receiving most practice. (*d*) When corrected for practice effects, results were found to be compatible with the hypothesis that warm-up is positively related to amount of prerest practice.

REFERENCES

ADAMS, J. A., & REYNOLDS, B. Effect of shift in distribution of practice conditions following interpolated rest. *J. exp. Psychol.*, 1954, **47**, 32–36.

AMMONS, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.

BELL, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**(1, Whole No. 243).

BUXTON, C. E. Level of mastery and reminiscence in pursuit learning. *J. exp. Psychol.*, 1943, **32**, 176–180.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

IRION, A. L. Reminiscence in pursuit-rotor learning as a function of rest and of amount of prerest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.

JAHNKE, J. C. Retention in motor learning as a function of amount of practice and rest. *J. exp. Psychol.*, 1958, **55**, 270–273.

JAHNKE, J. C., & DUNCAN, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52**, 273–282.

KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.

KIMBLE, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.

LEWIS, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

WASSERMAN, H. N. The effect of motivation and amount of prerest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

(Received September 13, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 612–617

THE CS-UCS INTERVAL IN CONDITIONING SHORT- AND LONG-LATENCY RESPONSES[1]

JOAN E. JONES[2]

University of Sydney

The author has outlined a two-principle theory which predicts the effects of the CS-UCS interval in classical aversive conditioning (Jones, 1962). In brief, this theory is based upon the observation that variation in the CS-UCS interval affects the temporal relationship between (*a*) the CS and the UCR, and (*b*) the CR and the onset of the UCS. With the tentative assumption that UCS-onset is the reinforcing event, the effects of variations in the CS-UCS interval are treated in terms of the joint action of the principles of contiguity (CS-UCR) and of reinforcement (CR-UCS). In both cases a bidirectional gradient is assumed to operate such that the greater the time interval between CS and UCR or CR and UCS the less effective the learning. It is assumed that the efficacy of conditioning at a given CS-UCS interval reflects the additive combination of the overlapping gradients of contiguity and reinforcement.

From this analysis it follows that any variation in experimental procedure which affects either or both of these temporal relationships will affect the function relating the CS-UCS interval to response strength. One such factor is the latency of the response which is being observed. Further, according to the definition given, reinforcement cannot have effect until the CR is evoked (Jones, 1962). Therefore, the contiguity of CS and UCR assumes prime importance in the initial establishing of an overt response. As the CR becomes stronger reinforcement assumes greater importance. This leads to the expectation that the interval function changes during the course of training, so that intervals of different lengths yield learning curves of different shapes.

Although the theory also makes testable predictions about backward conditioning, the present experiment is concerned with forward conditioning and was designed to test the following predictions: (*a*) When groups are trained with a short CS-UCS interval they perform relatively well initially but then show a decrement in performance while groups trained with longer intervals show a less rapid initial rise but perform relatively better later in training. (*b*) In conditioning a short-latency response the acquisition curve shows a relatively rapid initial rise. In the case of a long-latency response a higher asymptote is reached less rapidly. (*c*) Where a response has a relatively long latency and there is some tendency for the CS to elicit that response from the outset a double-peaked function relating the CS-UCS interval to response strength will be obtained. The longer the latency of the response, the greater the separation of the peaks. (*d*) When conditioning responses which differ in their characteristic latency the interval function for the longer latency response is less sharply peaked and a wider range of intervals is effective.

---

[1] Based on a thesis submitted in partial fulfilment of the requirements for the PhD degree, University of Sydney, 1959. This research was supervised by R. A. Champion and part of the apparatus used in this study was provided by Commonwealth of Australia Research Grant No. 2002. Appreciation is expressed to R. B. Bromiley and P. J. Foley for their helpful criticisms of drafts of this article.

[2] Now at the Defence Research Medical Laboratories, Toronto, Canada.

METHOD

The responses selected were finger movement and the galvanic skin response (GSR). As measured in this experiment, the finger movement had a characteristic latency of 200–250 msec. in response to shock (UCS by instruction) and of 300–350 msec. to a tone (CS). The mean latency of the GSR was 1500 msec. in response to shock (UCS) and 2000 msec. in response to tone (CS). Both responses were conditioned at the same time in each *S* to ensure that they were acquired under identical conditions. Seven CS-UCS intervals were sampled.

Subjects.—The *S*s were 126 volunteers from an introductory course in psychology who were naive regarding the conditioning technique.

Apparatus.—The room containing *E* and the main apparatus adjoined the test room which was lined with a sound reducing material. The test room was well illuminated and circulation of air was maintained by an electric fan. The *S* sat in a comfortable chair equipped with arm and foot rests. The CS was a 2000-cycle, 75-db. tone of 30-msec. duration produced by an electronic oscillator. It was presented to *S* through earphones which at the same time muffled other noises. The UCS was a spike wave ac shock of 24 v., 1.60 ma., and 90 cps of 500-msec. duration, also produced by an electronic oscillator. The shock levels were measured with a 15,000 ohm resistor in place of *S*. The shock was delivered to *S* by means of copper electrodes in the form of spring clamps 15 mm. wide which almost encircled the second and the fourth fingers of the right hand. The duration of CS, UCS, and interonset interval were controlled by electronic timers. On the first and the third fingers of the right hand *S* wore thimble-type electrodes of German silver 30 mm. in length to allow the measurement of the GSR. Electrode jelly was used to ensure good contact at all electrodes. The GSR was measured and amplified by an electronic dc amplifier of the bridge type supplying a constant current of 100 μa. to the electrodes. The four fingers of *S*'s right hand lay on a wooden rest which was depressed by the weight of the fingers. If the fingers were lifted at least 1 mm. a spring caused one contact to close, lighting a lamp on *E*'s control panel, and a second to open, stopping a chronoscope which had started with the presentation of the CS. A four-channel ink-writing recorder, run at 2 mm/sec, recorded the GSR and showed signals corresponding to CS, UCS, and finger movement.

Procedure.—The *S* was seated in the test room and informed of the nature of the stimuli to be presented. The only instruction was to raise the fingers of the right hand from the rest as quickly as possible whenever the shock was felt.

Two pretraining trials were administered to each *S*, the first of which was a UCS (shock) alone, and the second a CS (tone) alone trial. These served to familiarize *S* with the stimuli and showed whether either response was evoked by the CS before training. Twelve training trials followed. These comprised three blocks (a, b, and c) each consisting of three paired presentations of CS and UCS and one test trial (CS alone). The test trial was placed second, third, and fourth in Blocks a, b, and c, respectively, and the order of presentation of the blocks was varied for different *S*s to yield six orders in all. Each order was used for 3 *S*s in each treatment group of 18. The intertrial interval was at least 60 sec. and varied upward of this according to the time taken for *S*'s skin resistance to reach a steady level. The interval was rarely greater than 75 sec. and never exceeded 90 sec. Acquisition curves were obtained showing CR strength measured on Test Trial 1 after 1, 2, or 3 (an average of 2) reinforced trials, on Test Trial 2 after 4, 5, or 6 (average 5), and on Test Trial 3 after 7, 8, or 9 (average 8) reinforced trials. Seven groups were trained, one under each of the following CS-UCS intervals: 20 ± 5, 235, 440, 660, 860, 1045, and 1245 msec.

RESULTS

The method used for recording finger movement yielded only a presence/absence measure so that the index of the strength of the finger CR is the frequency of response evocation. The amplitude of the GSR was recorded as a change in conductance and scores were obtained by expressing these measures as a proportion of *S*'s pretraining response to the CS. This procedure allowed for

TABLE 1

ANALYSES OF VARIANCE OF GSR AMPLITUDES

Source	*df*	*MS*	*F*
CS-UCS interval	6	21.40	.73
Error (between *S*s)	119	29.40	
Amount of training	2	0.48	.26
Interval × Training	12	3.74	2.03*
Error (within *S*s)	238	1.84	
Total	377		

* $P < .05$.

individual differences in responsiveness and equated the groups in the pretraining period. A frequency measure was derived by taking scores greater than the median of those obtained from all 126 *S*s on the three test trials.

An analysis of variance of the GSR amplitude data, summarized in Table 1, yielded a significant interaction between interval and training ($P < .05$ and just greater than .01). Figure 1 shows that the character of this effect is not one of simple divergence of the learning curves but is one of difference in shape. The shorter-interval groups reach their maximum performance earlier in training than do the longer-interval groups. Group VI, however, does not show the improvement in Trial 3 which is characteristic of the curves for Groups IV, V, and VII. It is however, unlike the curves for the three shorter-interval groups which characteristically show a decline on later trials. Kendall's tau (τ) was calculated from a ranking of the seven groups in order of mean performance and an arbitrary ranking in order I to VII on each of the three test trials. The obtained values of τ were .33, .24, and −.41. The trend in the τ values may be taken as an objective index of the expected form of interaction, the theoretical expectation being that they will change from +1.00 to −1.00 over an extended period of training.

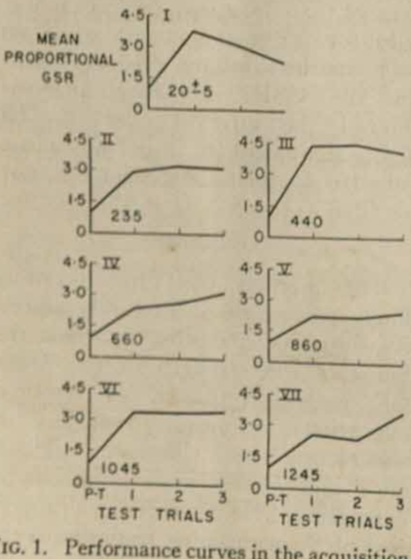


FIG. 1. Performance curves in the acquisition of the conditioned GSR.

TABLE 2

χ^2 ANALYSIS OF FREQUENCY OF FINGER CRs

Source	χ^2	*df*
CS-UCS interval (I) × Response occurrence (R)	12.21	6
Amount of training (T) × R	0.81	2
Response latency (L) × R	96.39**	1
I × T × R	10.35	12
I × L × R	13.32*	6
T × L × R	16.49**	2
I × T × L × R	7.68	12
Total	157.25**	41

* $P < .05$.
** $P < .001$.

The frequency data for finger movement were analyzed by partitioning of χ^2 (Sutcliffe, 1957). The Interval × Training interaction failed to reach the 5% level of significance but was of essentially the same kind

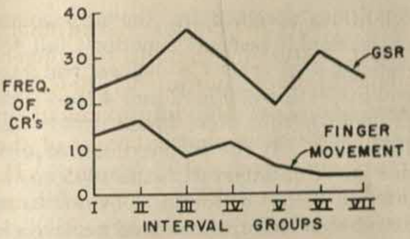


FIG. 2. Frequency of finger CRs as a function of CS-UCS interval.

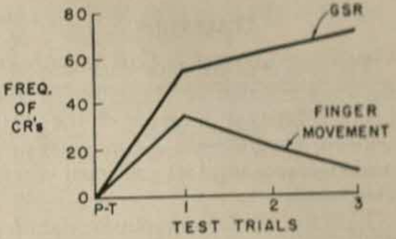


FIG. 3. The acquisition of conditioned GSR and finger movement.

as obtained in the GSR data, τ being .65, .41, and .24 on successive test trials.

Frequency data for the two responses were treated in one analysis which is summarized in Table 2. The Interval × Latency × Occurrence interaction was significant at the 5% level and the Training × Latency × Occurrence interaction at the .1% level. The former interaction, shown graphically in Fig. 2, indicates that the interval effect differs in the case of the two responses. The significance of this interaction may have been enhanced by summing over trials ("nonindependent" measures). However, the same effect appears in the analysis of the number of *S*s responding (below) which analysis does not suffer the same defect. The Training × Latency × Occurrence interaction (Fig. 3) indicates that the learning curves characteristic of the responses differ in shape. Many more conditioned GSRs were obtained than conditioned finger movements, as indicated by the Latency × Occurrence interaction which was significant at the .1% level.

TABLE 3

χ^2 ANALYSES OF THE NUMBER OF *S*s RESPONDING

Source	χ^2	*df*
CS-UCS interval (I) × Number of responders (N)	11.25	6
Response latency (L) × N	25.62**	1
I × L × N	16.68*	6
Total	53.55**	13

* $P < .05$.
** $P < .001$.

The frequency data were also compiled in the form of the number of *S*s responding on one or more test trials during the training period. Table 3 summarizes the analysis and the Interval × Latency × Responders interaction ($.01 < P < .02$) is shown in Fig. 4. Inspection of Fig. 4 indicates that the efficacy of intervals greater than 235 msec. declines markedly for finger movement while no such marked falling off is evident for the GSR. Separate analyses showed the interval effect to be highly significant for finger movement ($\chi^2 = 109.4$, $df = 6$, $P < .001$) but not for GSR ($\chi^2 = 5.2$, $df = 6$, $P \sim .50$).

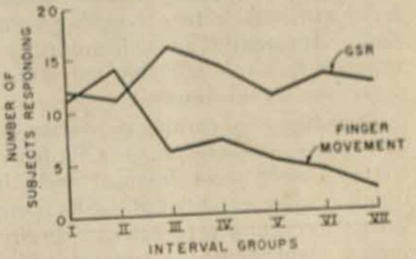


FIG. 4. Number of *S*s responding as a function of CS-UCS interval.

DISCUSSION

The results of the experiment support the interpretation of the effect of the CS-UCS interval in terms of the joint action of the gradients of contiguity and reinforcement. All four predicted effects were found.

Prediction a.—The amplitude data for the conditioned GSR showed that the shape of the learning curve was affected by the interstimulus interval used in training, short intervals giving a more rapid initial rise but a less well-sustained performance than longer intervals. A similar relationship appeared in the frequency data for finger movement but failed to satisfy the 5% level of confidence. This failure probably results from the insensitivity of the frequency measure. Further confirmation of Prediction a may be found in the literature. The curves presented by Moeller (1954) for GSR and McAllister (1953) for eyeblink show that longer-interval groups approach their performance asymptote at an initially slower rate than shorter-interval groups. Backward conditioning is treated by the theory as a case of true conditioning with a curve similar to that expected with very short forward intervals, i.e., showing a brief initial rise followed by a gradual decline. A curve such as this is described by Spooner and Kellogg (1947). In presenting the results above it was noted that Group VI, which was trained with an interval of 1045 msec., did not conform to neighboring groups conditioned with 660, 860, and 1245 msec. in that it showed no evidence of positive acceleration. This discrepancy does not appear to be attributable to a sampling variation. Apparently an unidentified extraneous factor is affecting the rate of acquisition at this interval.

Prediction b.—Learning curves for the two responses differed, finger CRs (short latency) being most frequent early in training while conditioned GSRs (long latency) continued to increase in frequency throughout the training period (Fig. 3).

Prediction c.—The GSR fulfils the conditions specified for the appearance of a double-peaked function (all *S*s responded to the CS before training) and inspection of Fig. 2 and 4 show that peaks or optima were found at 440 and 1045 msec. If the incongruous acquisition curve obtained with the 1045 msec.-interval should be found to be due to an extraneous factor, the second peak might be disturbed, at least early in training. Similar findings have been reported by Wickens, Gehman, and Sullivan (1959) who conditioned the GSR and obtained an optimum of 500 msec. with a second peak at 1500 msec. Other *E*s (Bierbaum, 1958; Moeller, 1954; White & Schlosberg, 1952) used too few intervals in the critical range to identify a second peak. In the present experiment a slight second peak in the function was obtained with finger movement. This was unanticipated as *S*s included in the study did not give a measurable response to the CS before training. A more sensitive measuring device may have detected a response in some *S*s. This is likely in view of the fact that *S*s were instructed to respond to the shock. The distance between the peaks (235 and 660 msec.) was less with finger movement than with GSR, in accordance with the prediction that the longer the response latency the greater the separation of the peaks when they occur.

Prediction d.—The functions relating CS-UCS interval to measures of CR strength differed as predicted, a more sharply peaked function being obtained for finger movement than for the GSR (Fig. 2 and 4), i.e., a narrower range of intervals was effective in conditioning the response of shorter latency.

The difficulties faced by other theorists in accounting for these effects in terms of the constructs they have proposed, namely, Hull's (1952) stimulus trace, Guthrie's (1935) proprioceptive stimulation, Hilgard and Marquis' (1940) reinforcement principle, and the grouping of stimulus events in cognitive-type theories (e.g., Tolman, 1932) have been discussed elsewhere (Jones, 1962). Briefly, none of these accounts can explain the obtained differences in interval functions for responses differing in their latency nor do they anticipate the changes in pattern among the curves of acquisition for the various intervals. The basic difference between these theories and the present one is that, while they rely on only one principle, either contiguity or reinforcement, in their account of the CS-UCS interval, this theory explains the effect in terms of the joint action of the two principles.

SUMMARY

An experiment was conducted in which a short- and a long-latency response were conditioned at the same time in seven groups of 18 *S*s, each group trained with a different CS-UCS interval, ranging from 20 to 1245 msec. The shapes of the learning curves obtained varied with the interval employed and the latency of the response being observed. The variations were consistent with predictions based on the two-principle theory of classical aversive conditioning. Differences between the functions relating performance to the CS-UCS interval for the two responses were also consistent with the theory.

REFERENCES

BIERBAUM, W. B. The temporal gradient in GSR conditioning. *J. gen. Psychol.*, 1958, **59**, 97–103.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

JONES, J. E. Contiguity and reinforcement in relation to CS-UCS intervals in conditioning. *Psychol. Rev.*, 1962, in press.

MCALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–423.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

SPOONER, A., & KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–334.

SUTCLIFFE, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychol. Bull.*, 1957, **54**, 134–137.

TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

WICKENS, D. D., GEHMAN, R. S., & SULLIVAN, S. N. The effect of differential onset time on the conditioned response strength to the elements of a stimulus complex. *J. exp. Psychol.*, 1959, **58**, 85–93.

(Received September 17, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 618–625

THE RETENTION OF INDIVIDUAL ITEMS[1]

BENNET B. MURDOCK, JR.

University of Vermont

The present paper reports a study of the short-term retention of individual verbal items. A study by Peterson and Peterson (1959) provided the starting point for the present experiments. In this study a single consonant syllable was presented to *S*, then immediately followed by a three-digit number. The *S* had to repeat the number, count backward by 3's or 4's at a 1-sec. rate for anywhere from 3 to 18 sec., then recall the original syllable. The results showed that the probability of recall decreased exponentially with duration of interpolated activity and approached an asymptote very close to zero.

Three experiments will be reported. Experiment I both replicated the original Peterson and Peterson (1959) study and investigated the effect of varying the nature of the to-be-remembered item. Experiment II was a study of short-term proactive inhibition (PI) which investigated the effect of varying the number of items preceding the to-be-remembered item. Experiment III studied the effect of varying the rate of the interpolated activity.

EXPERIMENT I

Method

This experiment was conducted in four successive sessions spaced about 4 weeks apart. Session 1 was a replication of the Peterson and Peterson (1959) study. The procedure was identical to theirs except for the following conditions: (*a*) the duration of the interpolated activity (i.e., counting backward) was 0, 3, 6, 9, 12, or 18 sec.; thus, we substituted the 0-sec. interval for the 15-sec. interval; (*b*) the metronome was set at rate of 1 beat/sec instead of 2 beats/sec; (*c*) the zero point of the interpolated activity was considered to begin with the first beat of the metronome following the presentation of the nonsense syllable; and (*d*) all *S*s were given sufficient preliminary practice on the interpolated activity to enable them to count backward accurately. As in the Peterson and Peterson (1959) study both latency and accuracy of response were recorded, the consonant syllables were selected from the same pool, and the same controls (counterbalancing of syllables among conditions and conditions among stages of practice) were employed.

Session 2 was an exact duplication of Session 1 except for the to-be-remembered items. Instead of consonant syllables the items were single, monosyllabic, nonhomophonic words selected from the Thorndike-Lorge (1944) list of the 1000 most common words in the English language.

To test whether the large differences obtained between the results of Sessions 1 and 2 could be attributed simply to the number of "chunks" of information (i.e., three unrelated letters in the first session or one common word in the second session; see Miller, 1956) Session 3 was an exact duplication of Sessions 1 and 2 except that the to-be-remembered items were word triads. That is, on each trial *S* was presented three monosyllabic nonhomophonic words from the same pool as above. In selecting the word triads the only stipulation was that the three words could be clearly spoken in the same length of time (i.e., 1 sec.) required to say the three letters of the consonant syllables of Session 1.

As both accuracy and latency of recall were used as dependent variables, Session 4 was conducted to determine whether the duration of the interpolated activity itself would affect response latency. No stimulus item other than the three-digit number was presented; *S* was merely required to say "stop" as quickly as possible after the interpolated activity (counting backward) was terminated.

The same *S*s were used in all four sessions;

---

[1] This study was supported by a research grant, M-3330, from the National Institutes of Health. Cynthia Marvin conducted many of the experiments and helped in the analysis of the results.

they were students of both sexes from the introductory psychology course who were fulfilling a course requirement. There were 24 *S*s in each of the first three sessions and 20 *S*s in the fourth session.

Results

The results of Session 1 clearly confirmed the results of Peterson and Peterson (1959). They presented their results in terms of the proportion of correct responses occurring with latencies less than 2.83 sec. (the overall mean latency for all correct responses). We analyzed our results in the same way and compared them with the values predicted by the formula given in Peterson and Peterson (1959, Fig. 3, p. 195). The differences between the results predicted by this formula and the results actually obtained in Session 1 were small and nonsystematic; the mean (algebraic) difference in proportion was −.025.

The mean proportion recalled at each retention interval is shown in Table 1. These results include all correct responses regardless of latency. To receive credit for a correct response in Sessions 1 and 3, *S* had to recall all three consonants or words in the correct order. Clearly, in all cases forgetting increased with the duration of interpolated activity, and with the longer retention intervals there was very little recall in Sessions 1 and 3.

TABLE 1

MEAN PROPORTION OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. I

Retention Interval	Session 1 CCC Trigrams		Session 2 Words		Session 3 Word Triads	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
0 sec.	.94	.11	.98	.04	.93	.10
3 sec.	.77	.15	.99	.03	.73	.21
6 sec.	.40	.18	.96	.07	.39	.23
9 sec.	.26	.19	.91	.10	.30	.21
12 sec.	.24	.14	.86	.11	.31	.22
18 sec.	.16	.12	.84	.14	.23	.16

As a simple statistical test, for each *S* we determined the total number of items recalled (maximum of 48) over all retention intervals. The mean values were 22.1, 44.3, and 23.0 items with *SD*s of 3.86, 1.99, and 5.98 for Sessions 1 to 3, respectively. The difference between Sessions 1 and 3 was not significant ($t = 0.94$), but the difference between Session 2 and the other two combined was highly significant ($t = 24.9$, $P < .001$).

An analysis of the latency data substantiated the above results. The mean latencies are shown in Table 2 and, in general, latencies increased with the duration of the retention interval. An analysis of variance of the results of Session 4 showed that

TABLE 2

MEAN LATENCY OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. I

Retention Interval	Session 1 CCC Trigrams		Session 2 Words		Session 3 Word Triads		Session 4 Simple RT	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
0 sec.	1.91	0.97	1.13	0.30	1.63	0.28	1.06	0.26
3 sec.	2.01	0.64	1.32	0.31	2.02	0.67	0.85	0.17
6 sec.	2.45	1.01	1.51	0.55	2.32	0.74	0.86	0.18
9 sec.	3.16	1.21	1.61	0.58	2.40	0.72	0.90	0.18
12 sec.	2.49	0.91	1.59	0.62	2.63	1.06	0.92	0.19
18 sec.	3.04	1.18	1.69	0.65	2.79	1.20	0.92	0.25

TABLE 3

MEAN PROPORTION OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. II

Retention Interval	Number of Prior Words					Mean
	0	3	6	9	12	
0 sec.	.995	1.000	.985	.985	.974	.988
3 sec.	1.000	.959	.927	.948	.969	.960
6 sec.	.953	.834	.896	.901	.865	.890
9 sec.	.938	.844	.860	.849	.875	.873
12 sec.	.938	.771	.834	.818	.860	.844
18 sec.	.907	.823	.802	.828	.823	.836
Mean	.955	.872	.884	.888	.894	

the duration of the interpolated activity did have a significant effect ($F = 25.14$, $P < .001$), but most of the difference was due to the 0-sec. retention interval. Thus, the "true" values for the 0-sec. interval for the first three sessions should be slightly lower than those reported. The corrections for differences among the remaining retention intervals are probably too slight to bother with.

To summarize the results of Exp. I: (*a*) they confirm the previous findings of Peterson and Peterson (1959); (*b*) they show that, for three types of items, forgetting, whether measured by accuracy or latency, increased with the duration of the interpolated activity; and (*c*) they show little difference between the retention of three consonants and three words but marked differences between the retention of three "chunks" and one word. The possible confounding of intersession practice effects will be considered with the results of the next experiment.

EXPERIMENT II

Method

This experiment was a study of short-term PI. As in Session 2 of Exp. I the stimulus items were single monosyllabic nonhomophonic words. A given item was preceded by 0, 3, 6, 9, or 12 words, also from the Thorndike-Lorge list of the 1,000 most common words. The preceding words were read at a 1-sec. rate, but always *S* was asked to recall only the stimulus item. The interpolated activity was, as in Exp. I, counting backward for 0, 3, 6, 9, 12, or 18 sec. Each *S* was tested under all conditions. The 5 × 6 design (five levels of prior words, six retention intervals) required 30 trials (lists) for one replication, and there were eight replications in all. Thus, each *S* was given a total of 240 trials, 60 on each of 4 days. There were 24 *S*s from the same population as before.

In all other respects the procedure was identical with that of Exp. I. There was complete counterbalancing of conditions within *S*s and stimulus material across *S*s. That is, there were eight blocks of 30 trials each so that the experimental conditions appeared equally often at all stages of practice. Within each block the order of the 30 trials was randomized. Each specific stimulus item was assigned equally often to each retention interval and, so far as possible, to each level of number of prior words. The specific words preceding the stimulus item were randomized by starting at a different place on the mimeographed word list for each *S*.

Results

An analysis of intersession practice effects showed that the proportion of words correctly recalled over the 4 days were .87, .87, .92, and .93, respectively. Although the magnitude of the improvement was rela-

TABLE 4

MEAN LATENCY OF CORRECT RECALLS AT EACH RETENTION INTERVAL: EXP. II

Retention Interval	Number of Prior Words					Mean
	0	3	6	9	12	
0 sec.	0.90	1.01	0.98	0.97	1.01	0.97
3 sec.	1.09	1.12	1.17	1.16	1.17	1.14
6 sec.	1.20	1.27	1.33	1.24	1.29	1.27
9 sec.	1.26	1.31	1.35	1.47	1.26	1.33
12 sec.	1.19	1.33	1.35	1.29	1.31	1.29
18 sec.	1.32	1.35	1.37	1.43	1.40	1.37
Mean	1.16	1.23	1.26	1.26	1.24	

tively slight, it was consistent enough to be highly significant ($F = 9.12$, $P < .001$). In Exp. I the proportion of correct recalls for the first session (consonant syllables) was .46, and the proportion of correct recalls for the third session (word triads) was .48. Since the slight improvement was about what would be expected by practice effects alone, the confounding of practice effects and stimulus items probably does not affect the conclusion that consonant syllables and word triads were equally well recalled.

The mean proportion of correct recalls, regardless of latency, at each retention interval in Exp. II is shown in Table 3, and the mean latency of correct recalls is shown in Table 4. Analyses of variance showed that, for each measure, both variables (number of prior words and retention interval) were significant at at least the .01 level. As the retention interval increased, accuracy decreased and latency increased. As the number of prior words increased, accuracy decreased from 0 to 3 words, then increased slightly but consistently from 3 to 12 words. As the number of prior words increased, latency increased from 0 to 3 words but then stayed relatively constant from 3 to 12 words.

When the control condition (0 prior words) is omitted, there were in all 264 extralist intrusions, 161 intralist intrusions, and 81 omissions. Thus, the number of extralist intrusions, intralist intrusions, and omissions occurred in approximately the ratio 3:2:1. An intralist intrusion was considered to be one of the words preceding the to-be-remembered stimulus item, an extralist intrusion was any other word, and an omission was a failure to respond. An analysis of the intralist intrusions showed that almost half (74, or 46%) of these intrusions were the word immediately preceding the to-be-remembered stimulus item, and in general the percentage decreased with increasing remoteness from the stimulus item.

If the control group is included, there were in all 324 extralist intrusions. Of these 324 intrusions 98 (30%) were words that had been correct on one of the preceding trials, 86 were classified as having been mispronounced by *S*, 10 were obvious associations or synonyms of the correct response, and 130 were unclassifiable. Of the 98 intrusions from previous trials 58 had been correct on the immediately preceding trial, 13 on the preceding trial once removed, and the number decreased regularly with increasing remoteness. Thus, in general the analysis of the errors showed marked recency effects for both intralist and extralist intrusions.

To summarize, the results of Exp. II showed that proactive inhibition does occur in the short-term retention of individual items but (at least for accuracy) appears to be a **U** shaped function of number of previous words.

EXPERIMENT III

Method

This experiment was a study of the effect of different rates of interpolated activity on short-term retention of individual items. For this purpose counting backward did not seem to be a suitable type of interpolation; *Ss* could not easily count faster than a 1-sec. rate, and a slower rate might allow rehearsal of the original item. Therefore, a new procedure was devised. A list of words of unknown length was read to *S*, and *S* was instructed to remember both the first and the last three words in the list. Requiring *S* to remember the last three words seemed to prevent rehearsal of the first word.

A preliminary experiment was conducted to test the feasibility of this method. There were lists of six different lengths: 4, 7, 11, 16, 21, and 31 words, all nonhomophonic words from the Thorndike-Lorge 1,000 most common. There were 10 counterbalanced replications of the six lengths for a total of 60 trials (or lists) in all. Group testing was used, and all words were read in time with an electric metronome beating at a rate of 1 beat/sec. After each list had been read *S*s had 20 sec. in which to write down the first and last three words of the list. A "ready" signal preceded by 2 sec. the start of each list. There were 29 *S*s from the same population as before, all tested in a single session.

An analysis of the results of this preliminary experiment showed that, with one reversal, the proportion of correct recalls of the first word decreased as list length increased. The minimum proportion was .841 for the 31 word list. Statistically, the effect of list length was highly significant ($F = 5.81$, $P < .001$) and, in general, the results were quite comparable to the analogous data on single words in Exp. I and II. Therefore, it was decided to use this method to study the effect of rate of interpolation on short-term retention.

The main experiment was conducted in three sessions and followed closely the procedure of the preliminary experiment. In Session 1 there were six list lengths: 4, 7, 10, 13, 16, and 19 words. The metronome was beating at a rate of 2 beats/sec; starting with the second beat after the first word had been read the remaining words were read either at every beat, every other beat, or every fourth beat. Thus, in effect, the first word was presented for 1 sec. and the remaining words were read at a rate of either 2, 1, or 0.5 words/sec. List length and presentation rate were orthogonal to each other, and there were three replications for a total of 54 trials (or lists). The order of presentation was randomized within each replication.

In Session 2 (actually requiring 2 days to complete) there were three list lengths (7, 13, and 19 words) but five different rates. The metronome was set at a rate of 2.5 beats/sec, and starting with the second beat the remaining words were read at every beat, every other beat, every third beat, every fourth beat, or every fifth beat. Thus, the rates used ranged from 0.5 to 2.5 words/sec. There were eight replications for a total of 120 trials, 60 on each of the 2 days. The order of presentation was randomized within each replication.

In Session 3 there were six list lengths (4, 7, 10, 13, 16, and 19 words) but only two different rates. The metronome was set at a rate of 2.0 beats/sec, and the entire list was read either at a rate of one word every beat or one word every third beat. Thus, the two rates were 2.0 and 0.67 words/sec. These two rates were selected on the basis of the results of Session 2 as those which should yield maximum differences. At the slower rate *S*s would have more time to rehearse the first word before the rest of the list started; this confounding was deliberately introduced in a further attempt to maximize the differences between the two rates. There were five replications for a total of 60 trials, and the order of presentation was randomized within each replication.

Group testing was used throughout. The same *S*s were tested in all three sessions (though not in the preliminary experiment). They were 27 students from the same population as before; however, only 21 were tested in Session 3.

Results

In each of the three sessions the proportion of correct recalls decreased as the retention interval increased. Analyses of variance showed that the effect was statistically significant at beyond the .001 level in all three cases.

In Session 1, the proportions of correct recalls were .827, .860, and .829 for the rates of 0.5, 1, and 2 words/sec, respectively. An analysis of variance showed that the differences were not statistically significant ($F < 1.00$). In Session 2, the proportions of correct recalls were .909, .940, .918, .963, and .946 for words read every 1, 2, 3, 4, and 5 beats, respectively. Again, the differences were not statistically significant ($F = 2.10$, $P > .05$). In Session 3, the proportions of correct recalls were almost identical for the two conditions (.894 and .895). Thus, there seems little doubt that, under the conditions of this experiment, the rate of interpolation is not an effective variable governing the short-term retention of individual items.

In Exp. III, *S*s were asked to recall

not only the first word but also the last three words. Since varying numbers of words preceded the last three words, an analysis of the recall data of the last three words as a function of number of preceding words can give further information about short-term proactive inhibition effects. The proportion of correct recall of the last three words is shown in Table 5. These results show exactly the same pattern as those of Exp. II (see Table 3): an initial drop followed by a gradual rise. The rise from 4 to 16 preceding words was reasonably linear, and the least-squares regression line of the means (bottom row of Table 5) had a slope significantly greater than zero ($t=4.76$, $df=3$, $P<.02$). Thus, these results confirm the conclusion of Exp. II; PI first increases but then decreases as the number of previous words increases.

An analysis of the errors showed omissions of 54%, 63%, and 64%; extralist intrusions of 28%, 25%, and 18%; and intralist intrusions of 18%, 12%, and 18%, each for Sessions 1, 2, and 3, respectively. Averaged across sessions the mean number of omissions, extralist intrusions, and intralist intrusions were 60%, 27%, and 12%, respectively. The corresponding values from Exp. II were 16%, 52%, and 32%, respectively. Thus, in Exp. III there were far more omissions and far fewer intrusions than in Exp. II. In Exp. II the interfering words preceded the to-be-remembered word while in Exp. III the interfering words followed the to-be-remembered word; this difference in procedure may have been responsible for the different ratios of omissions to intrusions found in the two experiments.

TABLE 5

Mean Proportion of Correct Recalls of the Last Three Words: Exp. III

Session	Number of Preceding Words					
	1	4	7	10	13	16
1	.938	.827	.846	.866	.863	.885
2	—	.874	—	.887	—	.898
3	.916	.867	.890	.895	.892	.905
Mean	.927	.856	.868	.883	.878	.896

TABLE 6

Mean Proportion of Correct Recalls of Each of the Last Three Words: Exp. III

Session	Word		
	Last—2	Last—1	Last
Prelim. exp.	.784	.914	.977
1st Session	.761	.884	.966
2nd Session	.770	.916	.974
3rd Session	.801	.906	.974
Mean	.779	.905	.973

There was a definite serial position effect in the recall of the last three words. Table 6 shows the proportion of correct recalls of the first, second, and third of these last three words (the third word being the last word in the list). The data show the typical serial position effect of free recall (Deese & Kaufman, 1957); there appears to be rather little interexperiment variability.

To summarize, the results of Exp. III showed no effect of rate of interpolation on short-term retention but confirmed the PI findings of Exp. II (and also see Rouse, 1959).

Discussion

The results of the present experiments substantiate and extend the previous word of Peterson and Peterson (1959). With an appropriate interpolated activity, the forgetting of an individual item will increase over a retention interval measured in seconds. Further, the interpolated activity evidently need not be highly similar to the to-be-remembered stimulus item; there would appear to be little formal similarity between the three-digit numbers used in counting backward and the consonant syllables, single words, and word triads used in Exp. I and II.

Given this low similarity, it would seem unlikely that the interpolated activity would interfere with the retention of the stimulus item in the same sense that, for instance, an A-C list interferes with the retention of an A-B list in paired-associate learning. However, the interpolated activity presumably prevents rehearsal of the item, and the results consistently showed that the more time that elapsed without rehearsal the poorer the retention. Thus, the results of the various experiments using the counting backward as an interpolated activity would seem to be more compatible with a decay theory than with an interference theory of forgetting.

On the other hand, the results of Exp. III (using, of course, a quite different interpolated activity) showed that the rate of interpolation was not a significant variable determining forgetting. Given a constant number of interfering items, the total time required to present these items (and hence the time for decay to occur) could vary over a range of three or four to one and still produce the same amount of forgetting. Since in Exp. III it was the number of interpolated items and not their rate of presentation that was the critical variable, these results are more compatible with an interference theory than with a decay theory. Thus, taken as a whole, the present experiments do not provide unambiguous support for either a decay theory or an interference theory of forgetting.

The curvilinear PI function found in Exp. II and III may, at least partly, reflect the "behavioral strategy" (Pollack, Johnson, & Knaff, 1959) adopted by *S*s. That is, in remembering the last word or words in a list of uncertain length, the more words that have been read the greater the probability that recall of the most recent word or words will be tested; *S*s can react accordingly. Although such a possibility cannot be completely ruled out it seems unlikely for two reasons: (*a*) As Table 5 shows, the PI function in Exp. III was essentially the same in Session 2 as in Sessions 1 and 3. However, the uncertainty about list length was only half as great in Session 2 and, as Pollack, Johnson, and Knaff (1959) suggest, behavioral strategy would vary with degree of certainty. (*b*) During the course of each session *S*s would gradually acquire a knowledge of the particular list lengths used, and this knowledge should facilitate the behavioral strategy. However, in Exp. III there was nothing approaching a significant practice effect, either in the preliminary experiment or in Sessions 1-3.

Finally, it is worth noting again the large differences in the retention of consonant syllables and word triads in one case, and individual words in the other case. Over the retention intervals studied, the proportion of correct recalls dropped to about .20 for the former and about .80 for the latter. It can be argued that the consonant syllables and word triads are three items to be remembered while an individual word is only a single item. If this is so, then, the number of items to be remembered would seem to be a significant variable in short-term retention.

SUMMARY

Three experiments were conducted on the short-term retention of individual items. Experiment I confirmed the findings of Peterson and Peterson (1959) and suggested that the number of items or "chunks" of information in the to-be-remembered stimulus item may be a significant variable in short-term retention. Experiment II demonstrated a significant proactive inhibition effect and suggested that the effect may be a U shaped function of number of preceding items. Experiment III showed that the rate of interpolated activity did not have a significant effect on short-term retention. The implications for a decay theory and an interference theory of forgetting were briefly discussed.

REFERENCES

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

ROUSE, R. O. Proactive inhibition as a function of degree of practice on the two tasks. *Amer. Psychologist*, 1959, **14**, 385. (Abstract)

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Bureau of Publications, 1944.

(Received September 21, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 626–627

SUPPLEMENTARY REPORT: PERCEPTION OF STIMULI OF VARYING DIMENSIONALITY

PAUL G. SHINKMAN[1]

IBM Research Center, Yorktown Heights, New York

Anderson and Fitts (1958) found that more information was transmitted by numerals superimposed on color patches than by either numerals or color patches alone, when these stimuli were presented for .1 sec. and *S* wrote down what he saw. With numbers alone, and colors alone, the stimuli had from 3 to 8 positions in a row, each containing one symbol from its corresponding alphabet; for the color-numeric stimuli there were from 2 to 4 color patches in a row, with a black numeral in the center of each patch. There were 9 colors, and 9 numerals (0 through 9 exclusive of 1). Generalizing their results, the authors concluded that increasing stimulus dimensionality (in this case combining numbers and colors into a single stimulus complex) improves information transmission.

Method.—The present experiment had three parts: Part I replicated a portion of the original study, and Parts II and III assessed the contribution of differences in visual angle to the main effect in the original work. In Part I three original conditions were repeated; 6 numbers in a row (Cond. N), 6 color patches in a row (Cond. C), and 3 numbers on 3 color patches (Cond. P). These conditions were chosen for replication because they yielded large between-condition differences in the original experiments. In Parts II and III two additional paired configurations of colors and numbers were used in which total visual angle was equated with that of the simple stimuli in Cond. N and C. One contained 6 elements in a row, with 3 color patches on the left and 3 numerals on the right (Cond. AP), and the other simply expanded the 3 superimposed pairs of Cond. P by putting in white spaces (Cond. EP), so that the visual angle for the complete stimulus was the same as for 6 elements in a row (2.68°).

Five *Ss* were tested in six sessions each for Part I. In each session 81 stimuli occurred for each of Cond. N, C, and P. The three conditions were counterbalanced over *Ss* and sessions. Within conditions, the stimuli were presented in a different random order for each session. In Part II, *Ss* served for three sessions, in each of which 81 stimuli occurred for each of Cond. N, C, P, and AP. Part III also involved 5 *Ss* for three sessions each, this time under Cond. P and EP.

Other details of method were identical to those of Anderson and Fitts (1958), except as follows: (*a*) In the original study *Ss* viewed 2.5 × 3.2 cm. stimulus elements through a shutter device from an average distance of 11 ft. A Harvard tachistoscope was used here and *Ss* were 24 in. from elements $\frac{3}{16} \times \frac{1}{4}$ in. in size. Thus, the visual angle subtended by the present stimulus was within 8′ of arc of the original. (*b*) Anderson and Fitts provided 38 ft-c of illumination and a gray wash-out field, whereas the tachistoscope was illuminated by 10 ft-c, and had a pre- and postexposure field of brightness roughly equal to that of the exposure field. (*c*) The present experiment was *S*-paced with *E* recording *S*'s verbal responses; a left-to-right readout order was imposed on *S*. In the original study, *S* wrote down his responses, using whatever order suited him.

Results and discussion.—Table 1 shows information transmitted, by symbol position and condition. The results of Part I are similar to the original Anderson and Fitts data, although the present data show a more striking decrease in transmission from left to right, which may be a function of the imposed readout order. The increase in stimulus dimensionality brought about in the original work by combining colors and numbers involved the introduction of two symbol alphabets into one stimulus, and also a gain in compactness due to the superposition of symbols, making possible a smaller visual angle. Parts II and III of the present study indicate that this smaller angle was in part responsible for the superiority of the complex stimuli over the numbers as reported by Anderson and Fitts.

The present findings also raise two minor questions concerning statements made in the original study. (*a*) The colors were selected by Anderson and Fitts for equal discriminability, but present data on proportions of incorrect responses to the same 9 colors, summed over *Ss*, sessions, positions, and conditions, reveal large between-color differences in discriminability. (*b*) Separate informational analysis on each symbol position as

---

[1] Grateful acknowledgement is made to E. T. Klemmer, under whose direction this research was accomplished.

Now at the University of Michigan.

TABLE 1

Average Information Gained in Bits

Part	Condition	Symbol Position						Mean/ Symbol
		1	2	3	4	5	6	
I	Numerals	3.03	2.74	2.55	1.45	.90	1.37	2.01
	Colors	2.85	2.66	1.84	.89	.43	.40	1.51
	Pairs							
	Numerals	2.49	1.77	2.09				2.12
	Colors	2.99	2.86	2.65				2.83
II	Numerals	3.09	2.71	2.42	1.68	1.29	1.76	2.16
	Colors	2.98	2.91	2.19	1.29	.79	.67	1.80
	Pairs (Adjacent)							
	Colors/Numerals	2.84	2.60	2.52	1.95	1.44	1.58	2.16
	Pairs (Superimposed)							
	Numerals	2.65	2.26	2.43				2.45
	Colors	3.10	3.02	2.93				3.02
III	Pairs (Superimposed)							
	Numerals	2.88	2.31	2.43				2.54
	Colors	3.09	3.08	3.03				3.07
	Pairs (Expanded)							
	Numerals	1.80	2.37	1.37				1.85
	Colors	3.00	3.06	2.85				2.97

Note.—Averages based on data computed separately for each *S*. Part I: 5 *S*s, 486 responses per *S* per condition; Part II: 5 *S*s, 243 responses per *S* per condition; Part III: same as Part II.

done in the original work involves the assumption that errors are independent. For example, transposition errors, which actually transmit information, do not contribute to the obtained estimate of information transmission. On the basis of a pilot study in which transposition errors occurred with an only slightly greater than chance frequency, Anderson and Fitts argued that the corresponding bias in information estimation is negligible. In the present experiment, however, the obtained proportion of transposition errors among all adjacent pairs of errors was .124, compared with a chance expectancy of less than .016. This finding suggests that transposition errors and other positional bias effects are important in a perceptual task of this sort.

In general, the present results show that the Anderson and Fitts data are not only reproducible, but are also stable under minor procedural alterations, including changes in brightness level. Thus the conclusions of the original study are supported.

Reference

Anderson, N. S., & Fitts, P. M. Amount of information gained during brief exposures of numerals and colors. *J. exp. Psychol.*, 1958, **56**, 362–369.

(Received September 16, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 627–628

SUPPLEMENTARY REPORT: A SELECTION ARTIFACT IN ROCK'S STUDY OF THE ROLE OF REPETITION[1]

JOANNA P. WILLIAMS

Yale University

Rock (1957) found that *S*s who had the original items of a list of paired associates presented repeatedly on every trial learned no faster than *S*s for whom item repetition was omitted, i.e., where new pairs were substituted on every trial for those which *S* had answered incorrectly. He concluded that repetition did not aid learning. There is the possibility, however, that the items in the list were not equal in difficulty and that *S*s in the nonrepetition group eliminated the difficult items. If so, their final lists would be easier than those learned by the repetition *S*s. The effect of such a selection factor would oppose that of any possible effect of repetition. The present experiment replicates Rock's study, and in addition determines whether or not such a selection factor does operate in this situation.

---

[1] These data were collected while the writer was a Teaching Fellow at Harvard University. She is indebted to Fred D. Sheffield for guidance in planning and conducting this research.

TABLE 1

Number of Errors to Criterion: Part II

Measure	Group			
	I[a]	II	III	IV
Mean	20.6	21.1	15.2	28.4
SD	9.3	8.6	5.3	11.2

Note.—$N = 20$ in each group.

[a] In Part I, the mean for Group A ($N = 10$) was 19.4, $SD = 8.1$.

Method.—The *S*s were 100 students at the Harvard Summer School, all volunteers. An additional 14 *S*s were used, but were eliminated when it became clear that they would not reach criterion within a reasonable amount of time (3 from Groups I, II, and III, 2 from Groups A and IV, and 1 from Group B). The items were 50 randomly combined letter-number pairs. Each letter except *I* appeared as a single, and as a double letter; the numbers were from 1 to 50. Each item was exposed (manually) for 3 sec., with 2 sec. between successive exposures. Training trials, on which both letters and numbers were presented, were alternated with test trials, on which each letter was presented for 5 sec. (with no additional interval between cards) and *S*'s task was to give the appropriate number. No knowledge of results was given during test trials. A 30-sec. interval followed both training and test trials. The criterion was one errorless trial. The procedure was demonstrated by giving *S* one training and one test trial on a list of six word pairs. The *S*s for whom there was to be substitution of pairs were told to expect such substitution.

In Part I, 20 *S*s were assigned randomly to one of two conditions. A list of 12 items, chosen at random from the 50-item pool, was presented to each *S*. Group A had Rock's nonrepetition procedure, in which new items were substituted on each trial for those items which *S* had failed to respond to correctly on the previous trial. In Group B, items which were answered *correctly* were eliminated from the list on the next trial, and new ones substituted. (For these *S*s, the experiment was continued until all 50 items had been presented.) A "final list" was made up from each of these 20 *S*s. The list from each *S* in Group A was simply that list on which the criterion was reached. The final list from each *S* in Group B was composed of the 12 items on which he had made the greatest number of errors.

In Part II, 80 *S*s were assigned randomly to one of four groups. A list of 12 items was randomly selected from the item pool for each *S* of Groups I and II. The *S*s in Group I learned their lists under the same nonrepetition conditions as Group A. Group II learned their lists with repetition of the original items on each trial. These two groups constitute a replication of Rock's experiment. Groups III and IV learned lists under repetition conditions like those of Group II, but their lists were preselected. Each of the first 10 *S*s in Group III learned one of the final lists from Group A; each of the second 10 learned a final list taken from an *S* of the identically treated Group I. Each final list obtained from Group B was presented to 2 *S*s in Group IV.

Results.—The mean number of failures to make correct responses before criterion for each group in Part II is presented in Table 1; the overall $F = 7.60$, $df = 3/76$, $P < .001$. No significant difference was found between Groups I and II, thus replicating Rock's original finding. The crucial comparison is between Groups II and III: Group III made significantly fewer errors than did Group II ($P < .05$), indicating that the lists learned by Rock's nonrepetition group may well have been easier than those learned by his repetition group. Also, Group IV made a significantly greater number of errors than did Group II ($P < .01$), showing that items could be selected which were more difficult to learn than a random list.

One important source of selection was noted: items which included a low number (1–15) were more likely to be found in the lists on which *S*s reached criterion than items with numbers 16–50 ($\chi^2 = 7.77$, $df = 1$, $P < .01$). Items composed of a low number were less likely to appear in those lists found to be more difficult to learn than random lists ($\chi^2 = 14.33$, $df = 1$, $P < .01$).

These results suggest that there may indeed be an important selection bias in the Rock experiment, which would be expected to enhance the performance of the nonrepetition group and to balance any possible effects of the repetition received by the other group. Thus Rock's results cannot be interpreted as evidence for all-or-none learning. There is probably a still greater amount of selection in the Rock procedure than is demonstrated here, because of the likelihood of selection on the basis of idiosyncratic differences in difficulty.

References

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, 70, [illegible]-193.

(Received October 31, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 629

SUPPLEMENTARY REPORT: ALTERNATION IN A T MAZE WITH HABIT HELD CONSTANT[1]

MERRELL E. THOMPSON AND EDWIN J. MARTIN

New Mexico State University

In an earlier study (Thompson, 1960) designed to check the role of stimulus, reinforcement, and choice conditions in alternation behavior, and to explore the possible interaction of these variables, some tentative conclusions were drawn concerning the importance of reinforcement and the interaction of the stimulus variable with conditions of reinforcement. Although indirect verification was also found for a number of deductions from the senior author's (Thompson, 1960), two-factor theory of inhibition, certain factors in the design made for difficulty in rigorous interpretation of the results. For example, an attempt to check Glanzer's (1953) study unfortunately contaminated the data by use of both constant and reversed stimuli. The influence of the spaced trials upon massed trial behavior was indeterminable, and habit to the various alternatives was uncontrolled (choice of the black or white alleys and right or left turns ranged from 2% to 98% for *S*s in the free choice groups and from 27% to 73% for the *S*s in the forced choice groups by the last day of the experiment).

Method.—The present study was designed to give a simple, more concise test of three of the deductions from the theory, as well as to check upon the earlier conclusions regarding the role of the reinforcement and stimulus variables and their interaction. It differed from the earlier study in the following respects: (*a*) all trials were given from opposite start boxes, thus presenting only reversed stimuli; (*b*) only massed trials were employed; (*c*) the first two trials each day (8 days) were free for all *S*s; and (*d*) habit to the various alternatives was kept constant by administering two forced trials, following the two daily free trials, in one of the following sequences: RRLL, LLRR, RLLR, and LRRL (expressed in terms of the stimulus component: WWBB, BBWW, WBBW, and BWWB).

The present study replicates the original in so far as (*a*) the two independent variables (stimulus conditions and reinforcement conditions) were manipulated at the same values; (*b*) the dependent variable of alternation was repeated; (*c*) Deductions 3, 4, and 5 from the two-factor theory were under test (Deductions 1, 2, and 6, tested in the earlier study, are undergoing further study in two additional investigations); and (*d*) the apparatus and procedure remained unchanged, except as described above. The *S*s were 80 rats of the Holtzman strain, 70–90 days old.

TABLE 1

MEAN NUMBER OF RESPONSE ALTERNATIONS

Group	Conditions	Mean	t[a]
I	B-B +	3.15	2.18*
II	B-B ○	5.35	4.09**
III	B-W +	3.35	1.81
IV	B-W ○	3.30	2.19*

[a] Tests were made comparing mean number of alternations with chance expection.
* $P < .05$.
** $P < .001$.

Results.—Table 1, which shows the mean number of response alternations for the four groups over the 8 days, reveals that Group II alternated responses significantly above chance. Groups I and IV showed response alternation significantly below chance, i.e., they alternated stimuli significantly above chance.

A three-dimensional analysis of variance, using the four replications as the third variable was performed after Bartlett's test had shown no significant heterogeneity of variance. As in the earlier study, the main effect of reinforcement condition was significant ($F = 10.27$, $df = 1/3$, $P < .05$). Again also stimulus condition was not significant as a main effect, but it interacted with conditions of reinforcement ($F = 32.46$, $df = 1/3$, $P < .025$). In t tests between various groups, significant differences occurred only when Group II was compared with some other group.

As in the 1960 study, Deductions 4 (that nonrewarded *S*s will alternate stimuli and repeat responses in a Cross T maze when the maze arms are dissimilar) and 5 (that nonrewarded *S*s will alternate responses and repeat stimuli in a Cross T maze when the arms are similar) were verified. Support was not found for Deduction 3 (rewarded *S*s will alternate responses and repeat stimuli in the Cross T maze).

REFERENCES

GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.

THOMPSON, M. E. Alternation in a T maze as a function of three variables. *Psychol. Rep.*, 1960, **7**, 103–110.

(Received November 4, 1960)

---

[1] This work was supported by Grant G-5817 from the National Science Foundation and was reported in a paper given at the 1960 meeting of the Rocky Mountain Psychological Association by the junior author.

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 630

SUPPLEMENTARY REPORT: UNLEARNING THE GAMBLER'S FALLACY[1]

HAROLD LINDMAN AND WARD EDWARDS

Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan

In a 1,000-trial probability learning experiment, Edwards (1961) found that as the length of a homogeneous outcome run increases, the probability that *S* will predict that event increases; this might be called a positive recency effect. Jarvik (1951) and Nicks (1959) have reported the opposite finding: as the length of a homogeneous outcome run increases, the probability that *S* will predict that event decreases. This is the gambler's fallacy. Jarvik's sequence had too few long outcome runs and so his *S*s' "gambler's fallacy" was not fallacious. But Nicks found a substantial gambler's fallacy even though he used a method of preparing sequences which makes every run statistic come out exactly at its expected frequency on the assumption of complete randomness and infinite sequence length. Edwards used the same sequence preparation procedure.

Edwards' (1961) 50–50 group data showed a small gambler's fallacy in the first 200 trials, no change of probability of prediction as a function of run length in the next 200 trials, and a positive recency effect in the last 600 trials. But even in the first 200 trials Edwards found less gambler's fallacy than did Nicks or Jarvik. Edwards and Nicks both used basic airmen at Lackland Air Force Base as *S*s. The disagreement between the Edwards and Nicks experiments indicates the desirability of a replication of the relevant part of Edwards' experiment using college student *S*s.

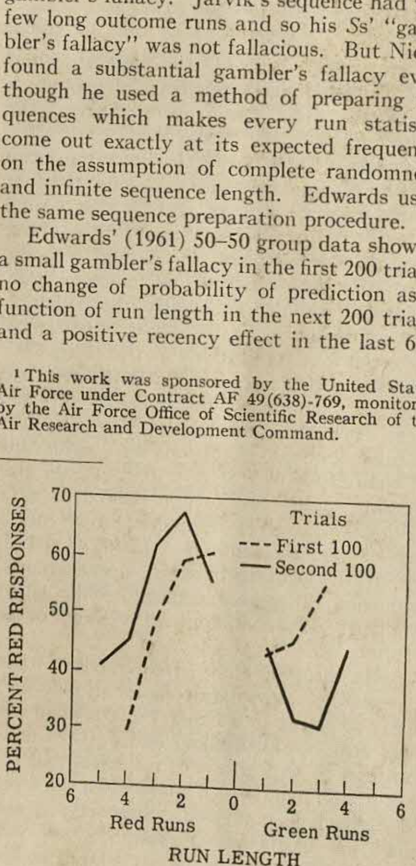


FIG. 1. Percentage of red responses following homogeneous outcome runs. (The *x* axis indicates the number of red or green outcomes included in the run. For example, the number 2 on the green runs side means that the number plotted is the percentage of times that a red response followed an outcome sequence consisting of a red followed by two greens. Runs which occurred less than three times per *S* in 100 trials are not plotted.)

Method.—Twenty-five male undergraduates were tested in groups of 5 to 8 *S*s. All *S*s saw the same sequence. Two hundred stimulus cards, 100 red and 100 green, were presented in a random sequence prepared by the method of constrained randomization used by Edwards (1961) and Nicks (1959). Before each card was presented, *S* predicted whether the card would be red or green by writing R or G on a small card, which was then placed face down in a box. The *S*s were told to maximize correct predictions; no information about sequence was given except that the cards were "well mixed."

Results.—As shown in Fig. 1, the negative recency effect over the entire 200 trials was slightly greater than that found previously by Edwards (1961) but slightly less than that found by Nicks (1959). The tendency found by Edwards for the gambler's fallacy to be replaced by a positive recency effect was confirmed in this study, with the additional finding that this process is already under way by Trial 100.

It is interesting to note that Feldman (1959), using a sequence with more long runs than would be expected by chance and college student *S*s, found no gambler's fallacy at all in 200 trials. The gambler's fallacy apparently is a very evanescent phenomenon—unless it is more often right than wrong, as it was in Jarvik's (1951) experiment.

REFERENCES

EDWARDS, W. Probability learning in 1,000 trials. *J. exp. Psychol.*, 1961, **62**, 381–390.

FELDMAN, J. On the negative recency hypothesis in the prediction of a series of binary symbols. *Amer. J. Psychol.*, 1959, **72**, 597–599.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

(Received January 26, 1961)

---

[1] This work was sponsored by the United States Air Force under Contract AF 49(638)-769, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 631

SUPPLEMENTARY REPORT: TWO-CHOICE DECISION BEHAVIOR WITH MANY ALTERNATIVE EVENTS

R. ALLEN GARDNER AND **JOHN B. FORSYTHE**

Wellesley College — *Bureau of the Census*

With multiple-choice versions of the familiar probability learning experiment it has been demonstrated that $P(A)$, the relative frequency of choices of A, the most frequent alternative, is an increasing function of the number of alternatives (Cotton & Rechtschaffen, 1958; Gardner, 1957, 1958). The present study was an attempt to reproduce this result by varying the number of alternative events in the series, $A, B, C, \ldots, N$, while restricting S to two choices, A or not-A.

Method.—There were two parallel and independent experiments. Within each experiment, $\pi(A)$ was held constant and the number of alternative events was varied. In Exp. I, $\pi(A)$ was .70 and there were four conditions having either 2, 3, 4, or 7 alternative events. In Exp. II, $\pi(A)$ was .60 and there were four conditions having either 2, 3, 5, or 8 alternatives. The Ss were 192 young soldiers drawn from the same population as those of two previous studies (Gardner, 1958, 1961). No S served in more than one condition and 24 served in each condition. The events consisted of block letters of the alphabet which were projected one at a time on the front wall of the experimental room. In all respects, the details of apparatus and procedure were identical to those of the two previous experiments with the following exception. Each response cabinet was provided with two telephone jacks and one plug. A small card was mounted over each jack. One card bore the letter designated as alternative A and the other was blank. The Ss were instructed to predict on each trial whether or not the letter designated as alternative A would be presented next by inserting the plug into the appropriate jack, either the lettered jack or the blank jack.

Results.—The obtained values of $P(A)$ were tabulated in three periods of 140 trials: Period 1, Trials 1–140; Period 2, Trials 141–280; and Period 3, Trials 281–420. These are shown in Table 1. In all of the statistical tests quoted there, the score for each S was taken as the number of his choices of alternative A during the period in question. For convenience, the means have been presented in the table as proportions over 140 trials. The analysis of variance performed on these data was the same as that described in Gardner (1958, 1961). The Fs for number of alternatives appear in the table and all are less than the 2.50 required for significance at the .05 level, most being less than 1.0. Moreover, the obtained values of $P(A)$ bear no orderly relationship to the number of alternatives in the series.

Under precisely the same conditions, but with the number of available choices kept equal to the number of alternative events in the sereis, $P(A)$ was a function of the number of alternatives (Gardner, 1958). In the present study, with S restricted to two choices, A or not-A, there was no detectable effect of the number of alternative events. Nevertheless, a variety of hypotheses remain open, for, when the number of choices was increased, and the number of alternative events held constant, the effect was distinctly different from that obtained when both the number of choices and alternative events were varied together (Gardner, 1961).

REFERENCES

COTTON, J. W., & RECHTSCHAFFEN, A. Replication Report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

GARDNER, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

GARDNER, R. A. Multiple-choice decision-behavior. *Amer. J. Psychol.*, 1958, **71**, 710–717.

GARDNER, R. A. Multiple-choice decision-behavior with dummy choices. *Amer. J. Psychol.*, 1961, **74**, 205–214.

(Received January 24, 1961)

TABLE 1

MEAN PROPORTIONAL FREQUENCY ($P[A]$) OF CHOICE OF ALTERNATIVE A IN THE THREE SETS OF 140 TRIALS IN EXP. I AND II

Period	Trials	Alternatives in Exp. I, $\pi(A) = .70$					Alternatives in Exp. II, $\pi(A) = .60$				
		2	3	4	7	F	2	3	5	8	F
1	1–140	.621	.635	.640	.669	0.93	.563	.614	.580	.557	1.73
2	141–280	.721	.738*	.748*	.757*	0.61	.611	.643	.601	.607	0.63
3	281–420	.764*	.751*	.735	.762*	0.47	.633	.685*	.618	.659*	1.74

* By t test computed on raw scores, $P(A) > \pi(A)$, significant beyond the .05 level with 23 *df* (two-tailed test).

Journal of Experimental Psychology
1961, Vol. 62, No. 6, 632

SUPPLEMENTARY REPORT: SHORT-TERM RETENTION AS A FUNCTION OF AVERAGE STORAGE LOAD[1]

KENNETH E. LLOYD

Washington State University

Lloyd, Reid, and Feallock (1960) and Reid, Lloyd, Brackett, and Hawkins (1961) have found recall errors to be systematically related to the average number of items *S*s were required to remember. The present study supplements this work by extending the *average storage load* to higher values, and by using as items words judged to be unrelated.

The *S*s listened to tape recorded sequences of items and recall points. An item consisted of a letter paired with a word; a recall point was indicated by the letter alone. The *S*s were instructed to recall all words previously paired with a given letter at each recall point. Between a given item and its recall point in the sequence other items, other recall points, or both, could be interpolated. A sequence contained 6 subsequences which varied in average storage load from 2.5 to 12.5. Each subsequence contained 20 items and 10 recall points. At each recall point either 1, 2, or 3 items were requested. All subsequences with a given average storage load contained four individual loads which bracketed that average, e.g., a 4.5 subsequence contained individual loads of 3, 4, 5, and 6. Six sequences were recorded to partially counterbalance all conditions.

The 64 words were arbitrarily grouped into eight letter classes. Items were recorded at a 4-sec. rate, and recall points at a 3-sec. rate. Each of the 12 *S*s, Washington State University undergraduates, served 5 days, and were reimbursed. They responded to two sequences on Days 1, 2, and 3, and to three sequences on Days 4 and 5.

The means of errors over the repeated measures for each day were averaged for all *S*s, and are shown in Table 1. While responding on Days 1 and 2, *S*s looked at the word list to aid learning the arbitrary classes; this procedure facilitated recall. A trend analysis of transformed scores on Day 5 resulted in a significant overall trend ($F = 227.07$, $df = 5/55$, $P < .01$) as well as significance for the linear, quadratic, cubic, and quartic components ($F\text{s} = 288.77, 157.75, 76.31$, and 12.00, respectively, $df\text{'s} = 1/11$, $P\text{s} < .01$). A second analysis of variance of raw scores for all 5 days yielded $P < .01$ for Load, Practice, and *S*s, and for the L × S and P × S interactions. Load and Practice did not interact.

Ideally errors for each *S* should increase with increases in load. With repeated measures the 12 *S*s provided 144 error curves. Inspection of adjacent curve points revealed 78 inversions out of 432 opportunities between the lower four load values. Between the higher three load values there were 139/288 inversions. This inversion probability of .48 suggests an asymptote at the higher loads.

In the three studies recall errors have increased regularly with average storage load over a wide range of its values, when related or unrelated items were used, and under conditions where *S*s recalled one or several items at each recall point.

REFERENCES

LLOYD, K. E., REID, L. S., & FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

REID, L. S., LLOYD, K. E., BRACKETT, H. R., & HAWKINS, W. F. Short-term retention as a function of average storage load and average load reduction. *J. exp. Psychol.*, 1961, **62**, 518–522.

(Received January 25, 1961)

---

[1] This research was supported by the United States Air Force through the United States Air Force Office of Scientific Research of ARDC under Contract No. AF 49(638)-805. This study is reported in greater detail in AFOSR Technical Note 60-1057, June 1960.

TABLE 1

MEAN ERRORS FOR AVERAGE STORAGE LOAD DURING 5 DAYS OF PRACTICE

Day	Average Storage Load											
	2.5		4.5		6.5		8.5		10.5		12.5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	3.08	2.47	9.96	3.79	12.37	2.83	13.87	1.95	13.37	2.21	14.12	1.98
2	1.92	1.51	8.08	2.56	11.08	2.73	12.29	2.41	11.50	2.56	13.50	2.0
3	3.46	1.69	10.83	2.68	12.67	3.57	14.42	2.88	14.71	2.10	15.58	2.7
4	2.37	1.52	8.97	2.15	12.31	2.58	13.97	2.88	13.62	1.82	14.37	2.62
5	1.52	1.46	8.12	2.92	10.61	2.65	12.01	2.38	12.26	2.47	13.70	2.41

LIBRARY